



# Lending Club Case Study

GROUP FACILITATOR – NIKITA SURVE

CONTRIBUTOR – SASHIKANT KUMAR

# Problem Statement

Lending club is a lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders.

Therefore, the objective of this case is to identify patterns or the factors which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.



# Data Exploration Steps:

This steps starts after researching about the data provided like understanding about each columns and the business objective. The steps included in data exploration is about getting to know the data, which includes.

- Understand the shape of the data. Here I had data with 111 columns and 39717 rows.
- See the data in each rows of the data to check if the first row of the table contains header name. Also, to see how the rows in each columns are present.
- See the datatype of each columns.
- See the basic statistics of numerical columns.

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_past_12m	pct_tl_nvr_dlt	percent_bc_gt_75	pub_rec_bank
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	NaN	NaN	NaN	NaN
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	NaN	NaN	NaN	NaN
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	NaN	NaN	NaN	NaN
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	NaN	NaN	NaN	NaN
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
39712	92187	92174	2500	2500	1075.0	36 months	8.07%	78.42	A	A4	...	NaN	NaN	NaN	NaN	NaN
39713	90665	90607	8500	8500	875.0	36 months	10.28%	275.38	C	C1	...	NaN	NaN	NaN	NaN	NaN
39714	90395	90390	5000	5000	1325.0	36 months	8.07%	156.84	A	A4	...	NaN	NaN	NaN	NaN	NaN
39715	90376	89243	5000	5000	650.0	36 months	7.43%	155.38	A	A2	...	NaN	NaN	NaN	NaN	NaN
39716	87023	86999	7500	7500	800.0	36 months	13.75%	255.43	E	E2	...	NaN	NaN	NaN	NaN	NaN

39717 rows x 111 columns

# Data Cleaning Steps:



Once we had the rough idea about the data. We start with the cleaning process. Which includes:

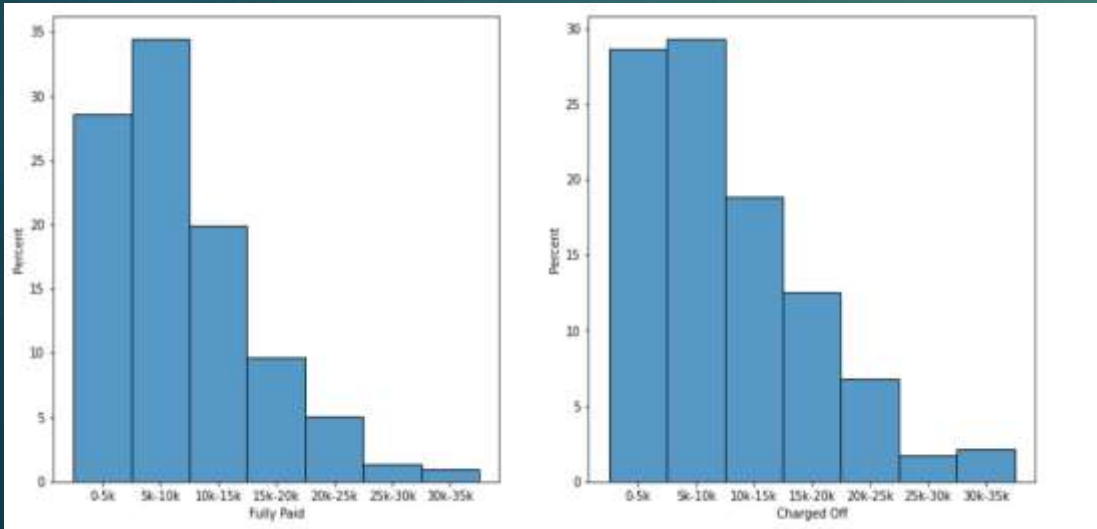
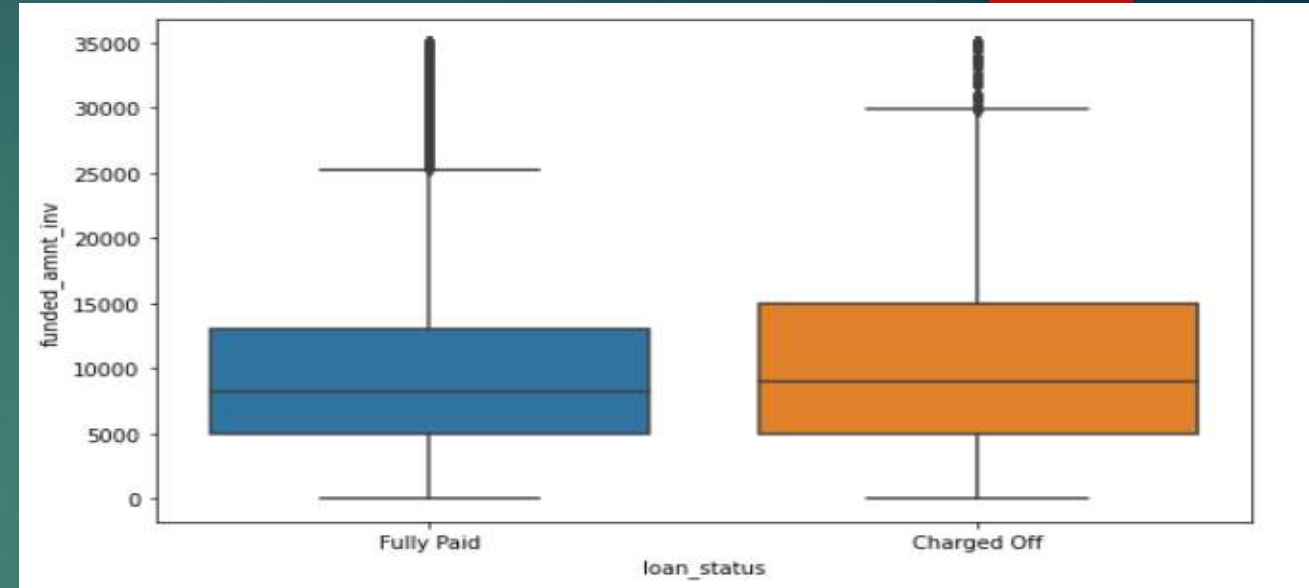
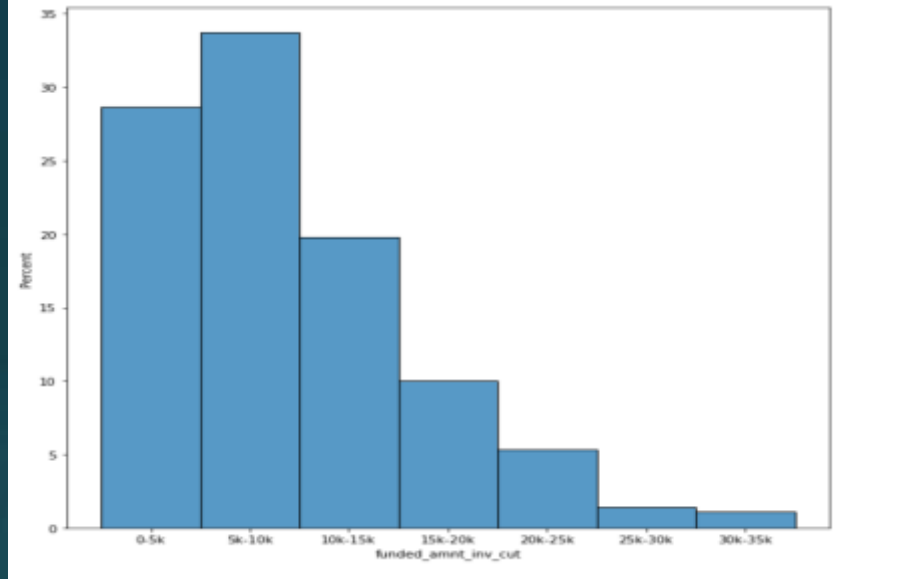
- Deleting columns which satisfied below criteria:
  - The columns with more than 30% of the missing values.
  - Columns with unique values or with incomplete values like pin code.
  - Columns which have only 1 unique value.
  - Text columns.
  - Columns related to post loan approvals.
- Converting the data into correct datatype.
- Split the datetime column into month and year.
- Filling the missing values. Using median for numerical and mode for categorical feature.
- Checking and removing outliers. Used percentile as a way to remove outliers.

# Columns Selected For Analysis.

- ▶ **Loan\_amnt**: the listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
- ▶ **Funded\_amnt**: the total amount committed to that loan at that point in time.
- ▶ **Funded\_amnt\_inv**: the total amount committed by investors for that loan at that point in time.
- ▶ **Term**: the number of payments on the loan. Values are in months and can be either 36 or 60.
- ▶ **Int\_rate**: interest rate on the loan
- ▶ **Installment**: the monthly payment owed by the borrower if the loan originates.
- ▶ **Grade**: lc assigned loan grade
- ▶ **Sub\_grade**: LC assigned loan subgrade
- ▶ **Emp\_length**: employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- ▶ **Home\_ownership**: the home ownership status provided by the borrower during registration
- ▶ **Annual\_inc**: the self-reported annual income provided by the borrower during registration.
- ▶ **Verification\_status**: indicates if income was verified by lc, not verified, or if the income source was verified
- ▶ **Loan\_status**: current status of the loan.
- ▶ **Purpose**: a category provided by the borrower for the loan request.
- ▶ **Dti(debt to income)**: a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested lc loan, divided by the borrower's self-reported monthly income.
- ▶ **Pub\_rec\_bankruptcies**: number of public record bankruptcies.
- ▶ **Issue\_d\_month**: the month which the loan was funded
- ▶ **Issue\_d\_yy**: the year which the loan was funded

# Univariate Segmented Analysis Of Numerical Features

## Funded amount invested by investors

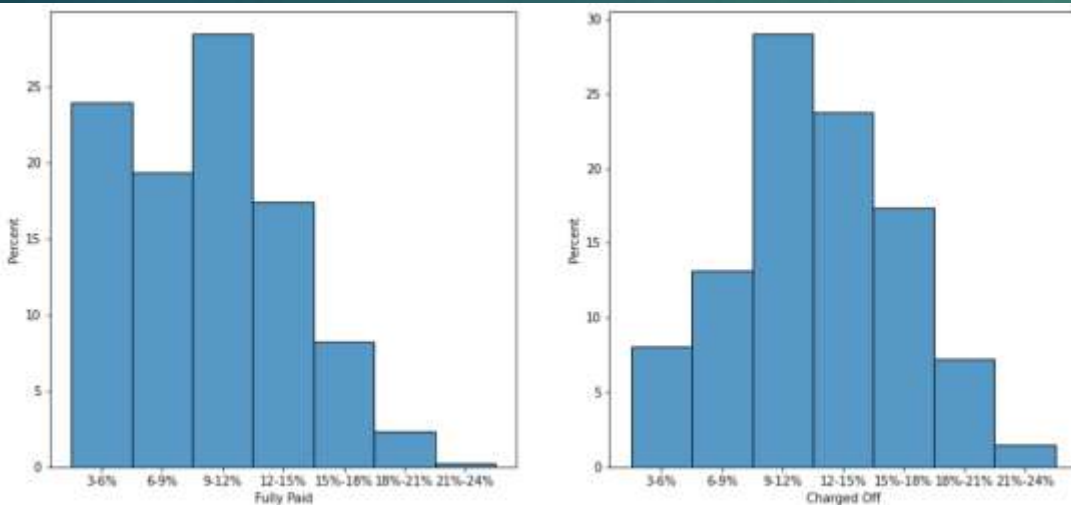
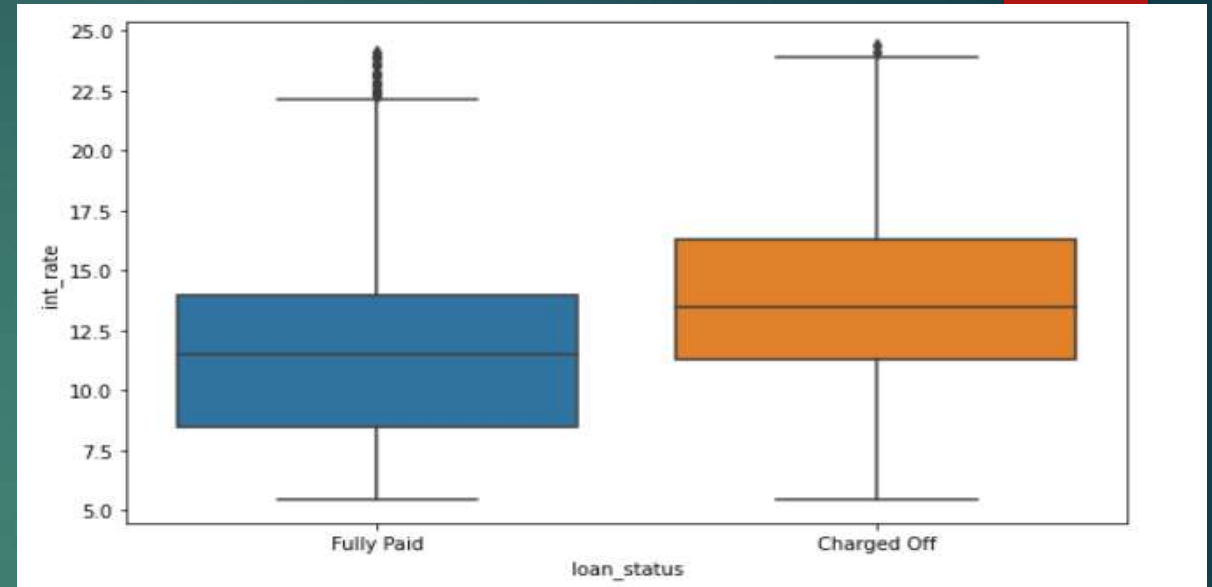
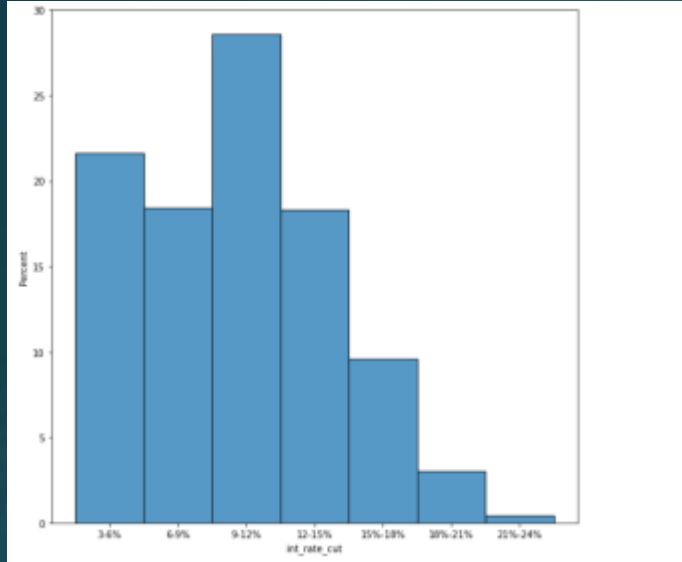


## Observation:

From the graph it is evident that more than up to 28% of the amount funded by the investor to the defaulter is between the 5k-10k and 27% of the funded amount is between 0-5k. However, for fully paid borrowers the highest funded amount is between 5k-10k with approx. 33% .

# Univariate Segmented Analysis Of Numerical Features

## Interest Rate



## Observation:

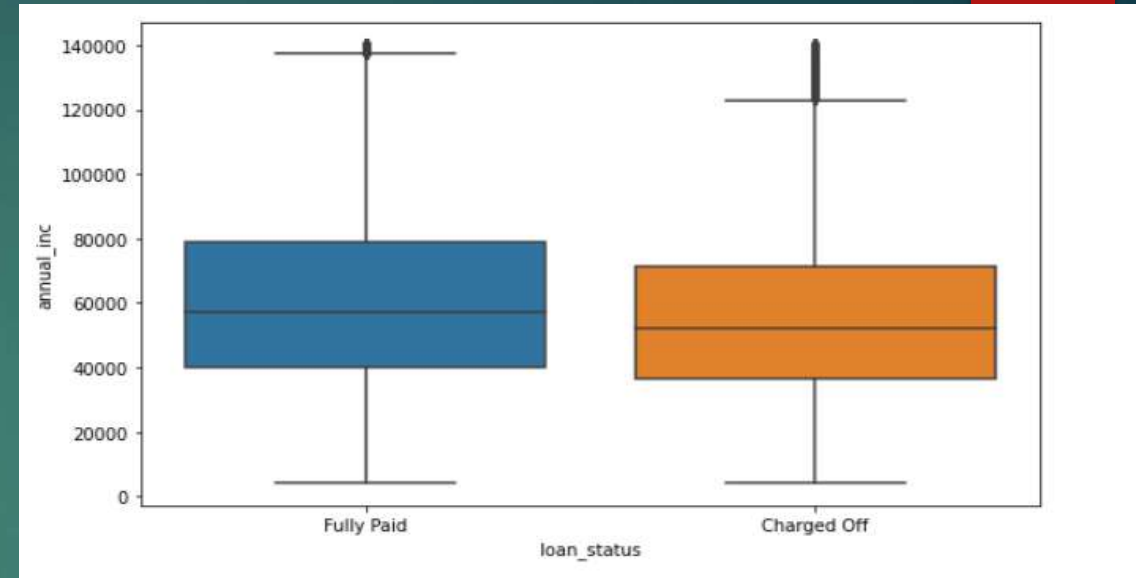
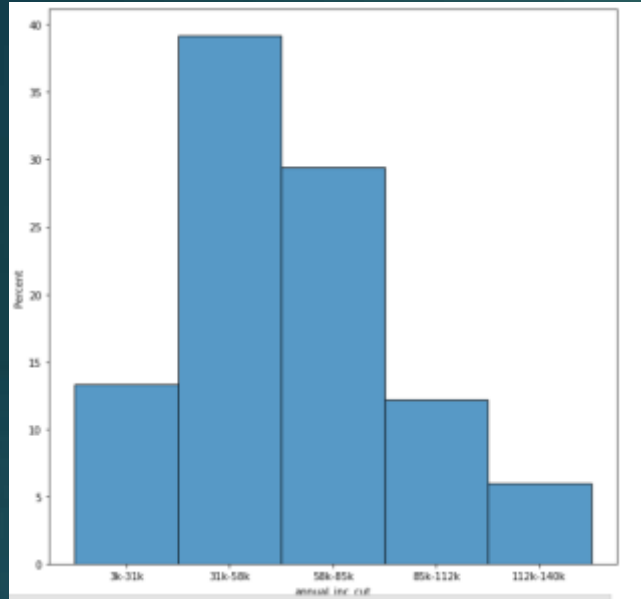
From the above analysis I can state that the maximum interest rate for fully paid borrowers and defaulters are between 9-12%.

However, it is also worth noticing that the around 24% of fully paid borrowers provided the loan with 3-6% interest rate whereas only approx. 7% of loan funded for defaulters with the same interest rate.



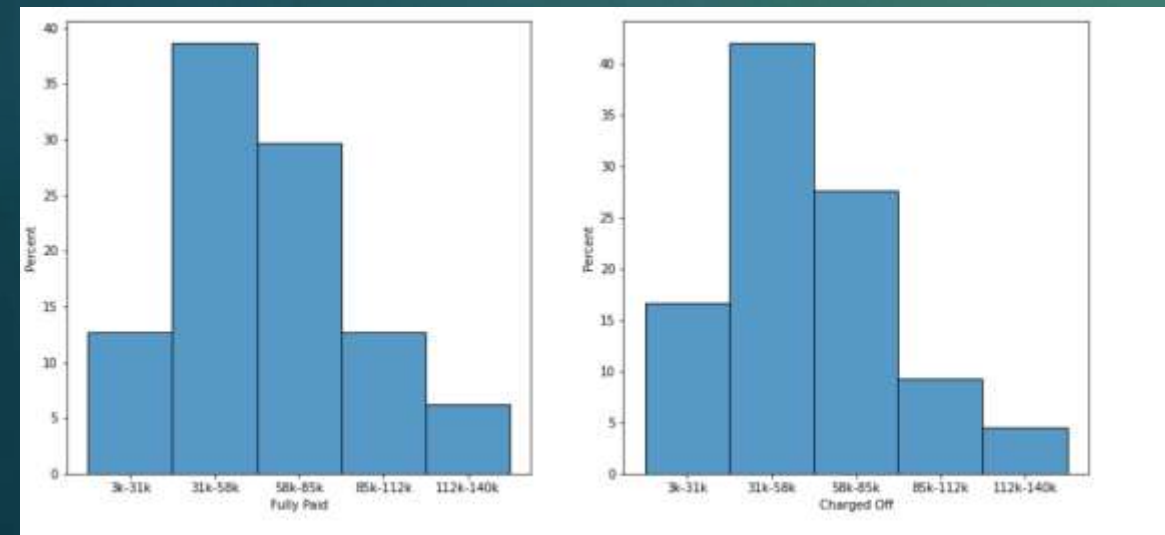
# Univariate Segmented Analysis Of Numerical Features

## ► Annual income:



## Observation:

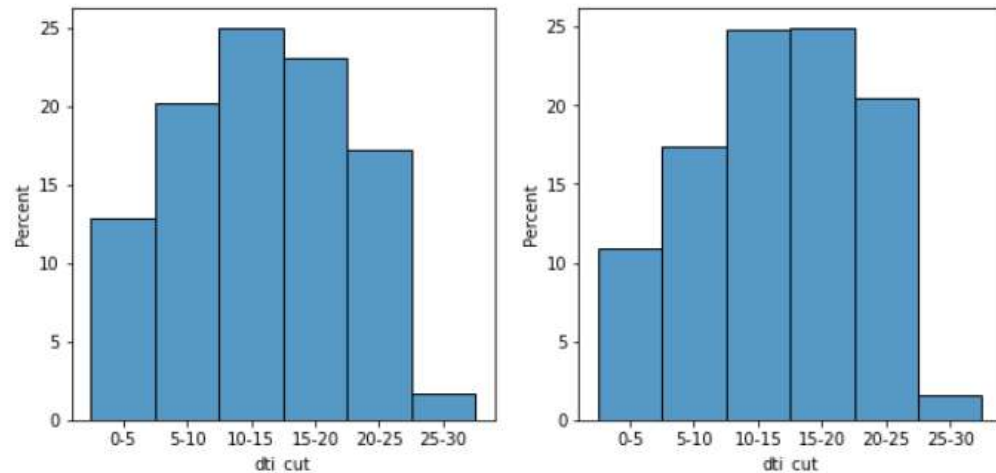
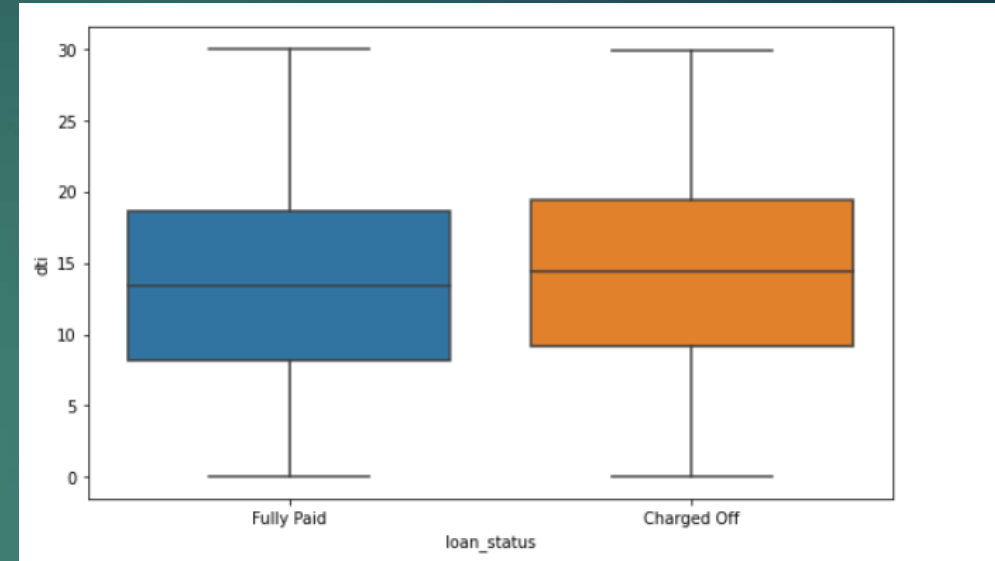
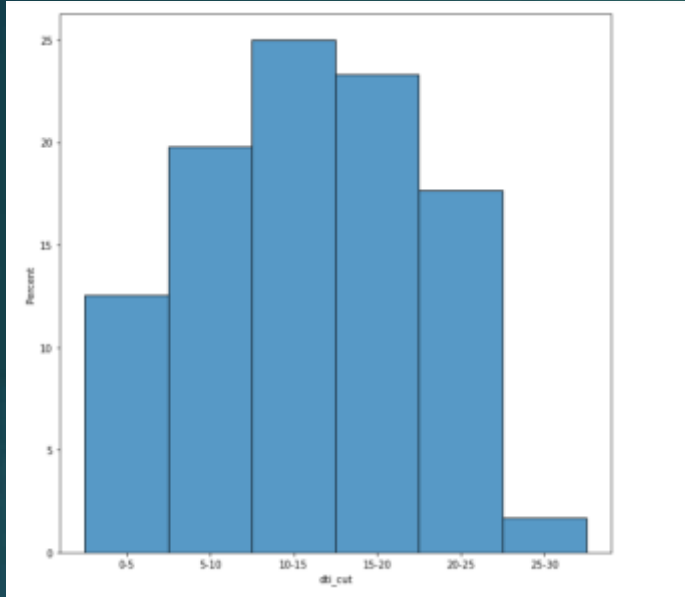
From the above analysis it is evident that for both loan status the highest percentage of borrowers who are defaulters or full payer of the loan have annual income between 31-58k.





# Univariate Segmented Analysis Of Numerical Features

## ► dti(debt to income):

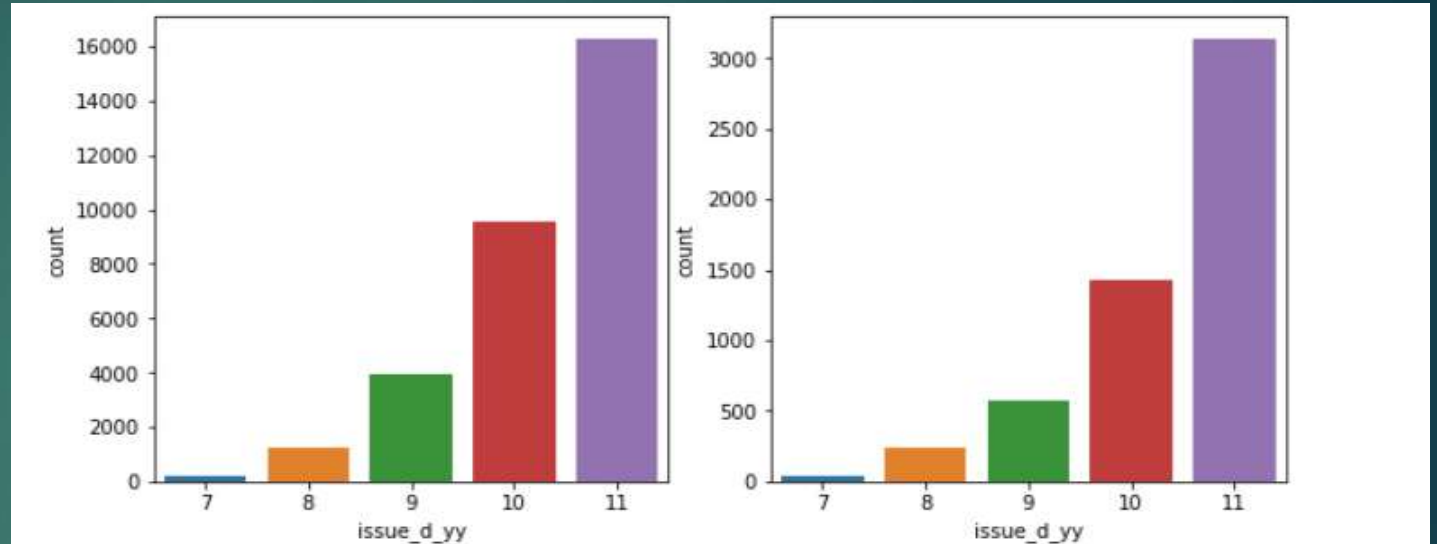
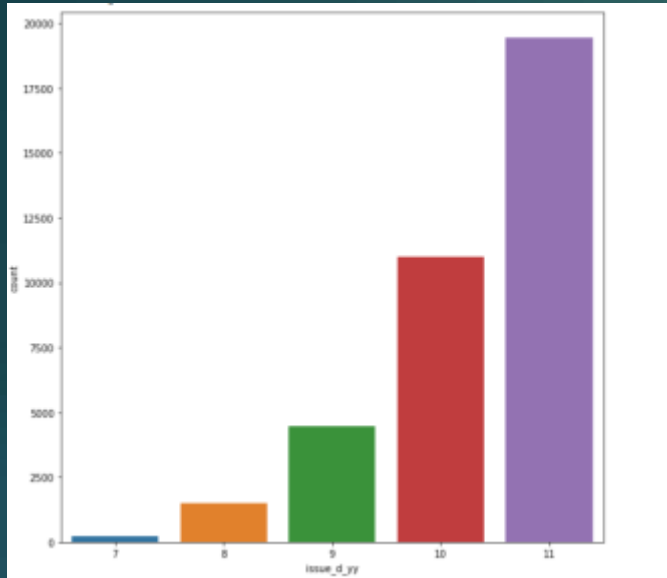


## Observation:

This ratio is highest between 10-15 for fully paid borrowers. But for charged off borrowers the highest ratio is between 10-20.

# Univariate Segmented Analysis Of Numerical Features

## ► Issue year:



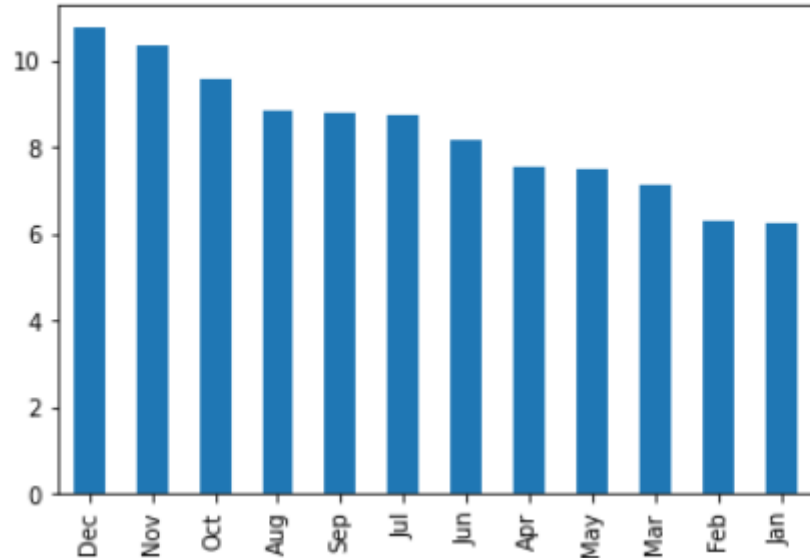
## Observation:

The maximum number of loan approved both for defaulter and Fully paid borrowers are in 2011.

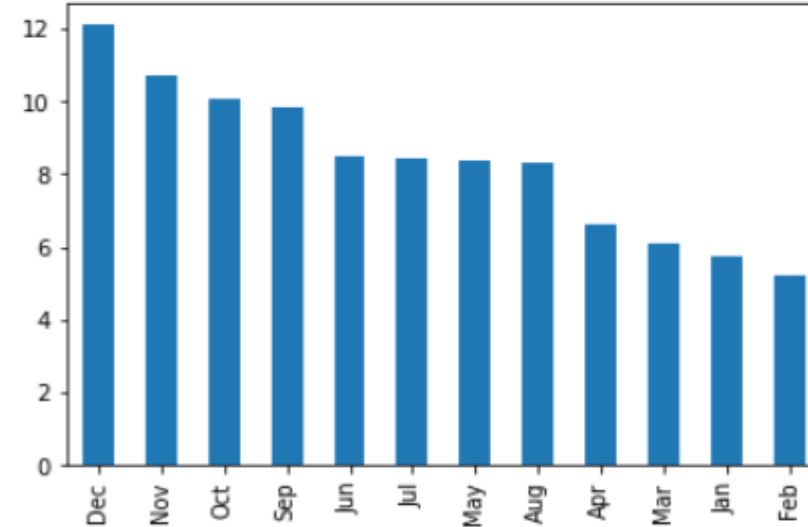
# Univariate Segmented Analysis Of Categorical Features

## ► Issue month:

Fully Paid



Charged Off

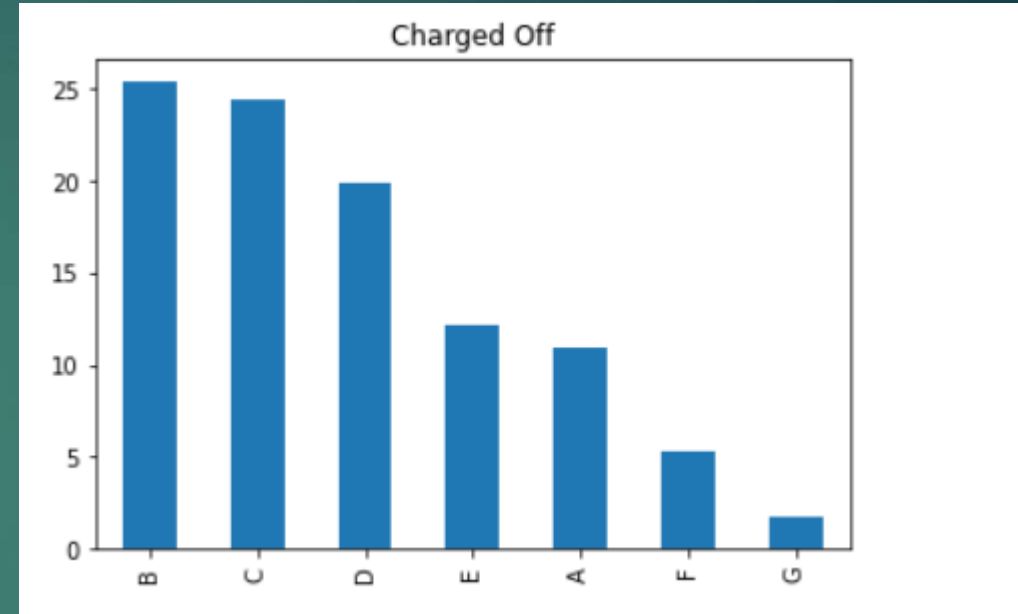
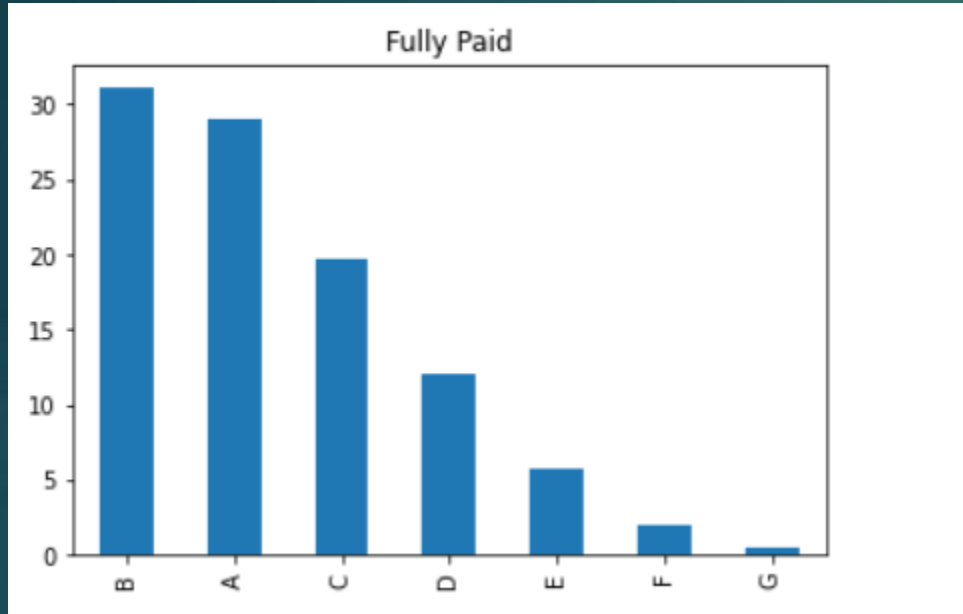


## Observation:

► From the analysis it is evident that maximum no. Of loans approved for both loan status is in the month of December.

# Univariate Segmented Analysis Of Categorical Features

## ► Grade:

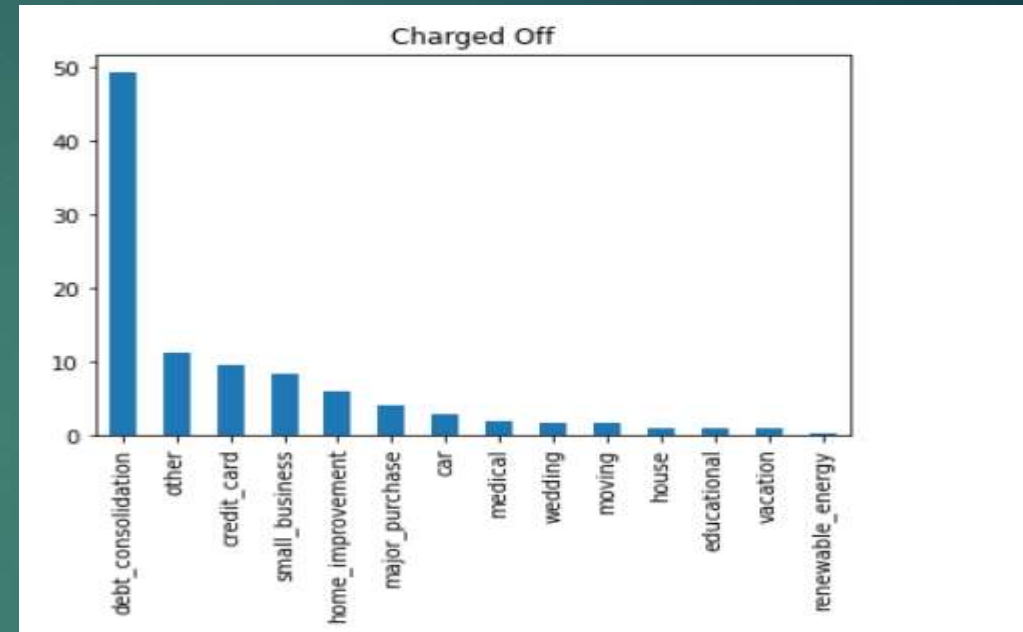
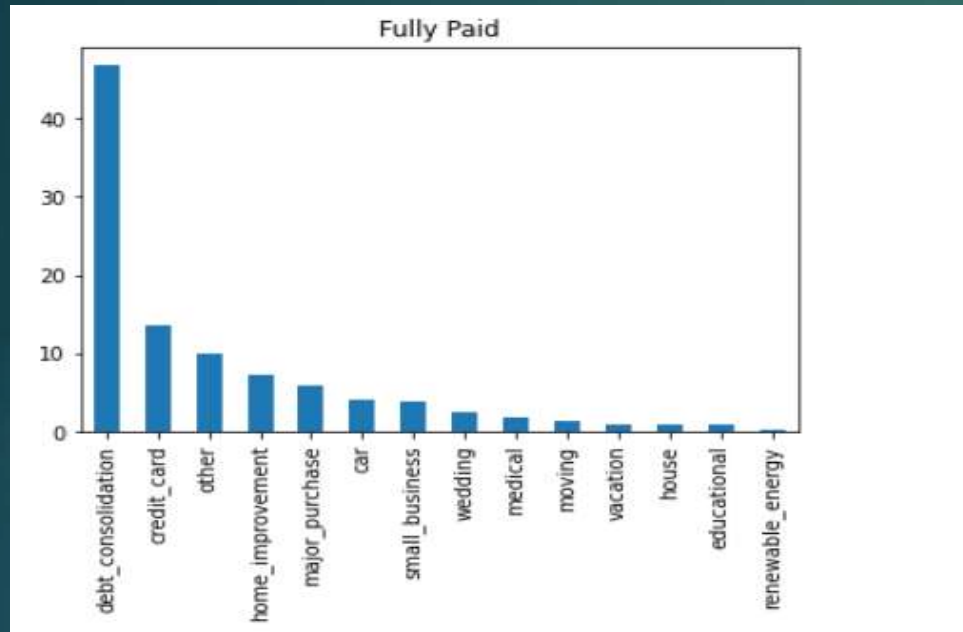


## Observation:

- From above graphs we can illustrate that majority of fully paid and defaulters belongs to grade B. But there is difference in the rest grades.
- From the bar plot for fully paid customer it is evident that around 28.66% of borrows belong to grade A and then 3rd highest borrower belongs to grade C. Whereas, for defaulters second highest grade is C with 23.80% followed by grade D with 19.8%.
- Borrowers belong to grade a is very less likely to be a defaulters.

# Univariate Segmented Analysis Of Categorical Features

## ► Purpose:

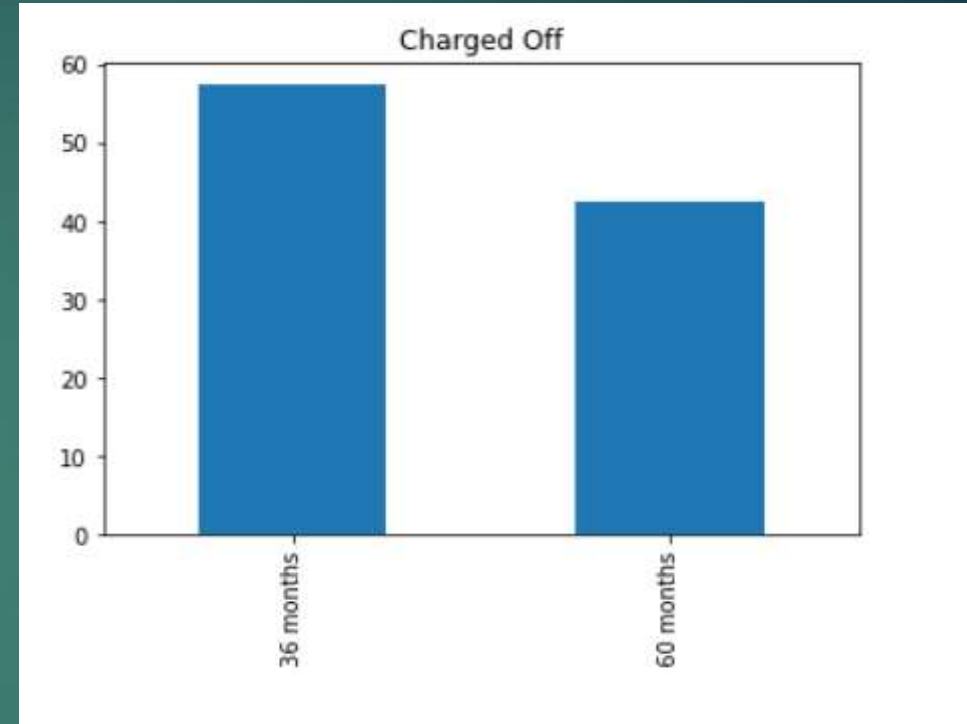
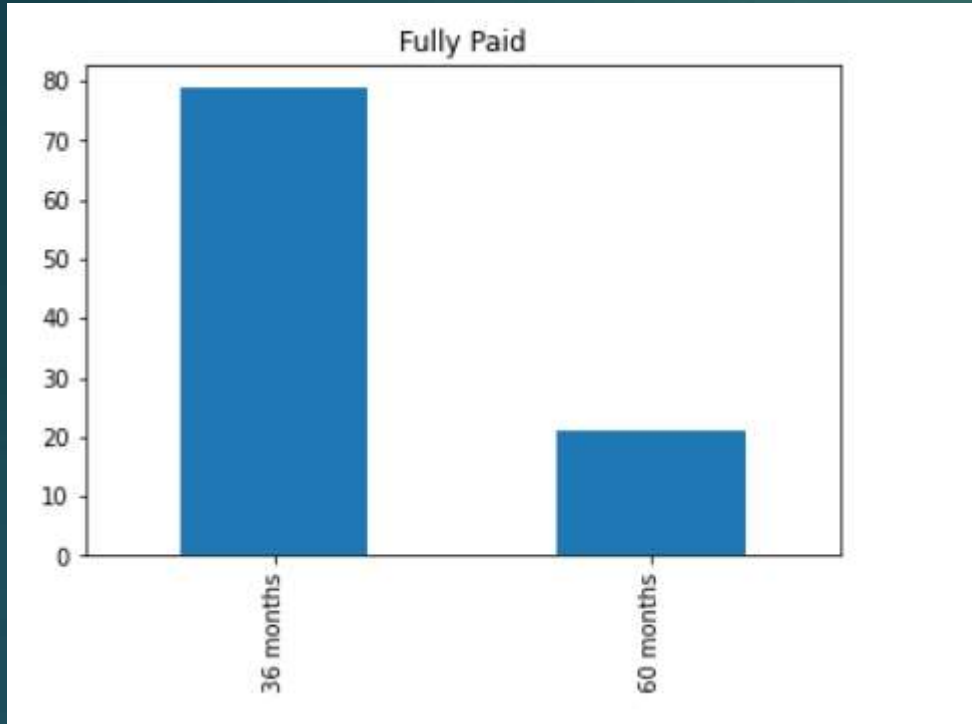


## Observation:

► From the above plot I can say that almost 50% of the time the loan taken by the charged off borrower is for debt\_consolidation.

# Univariate Segmented Analysis Of Categorical Features

## Term:



## Observation:

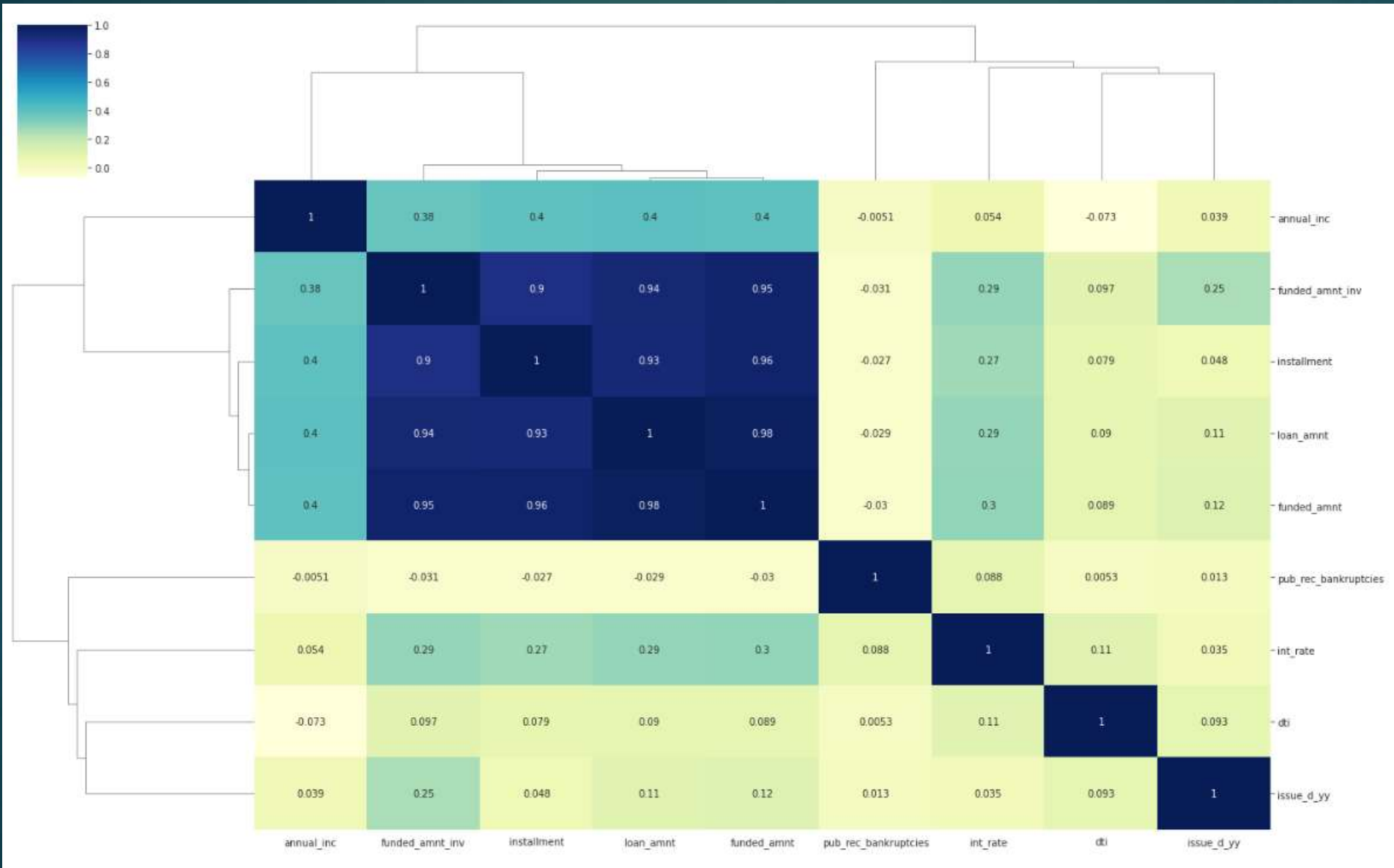
Both borrowers with loan status as fully paid and charged off prefer 36 months. But 40% of the defaulter prefer 60 months terms were as only 20% of the fully paid borrower goes for 60 months terms as the no of payment on the loan.

# Conclusion of segmented univariate Analysis

- ▶ If the funded amount invested by the investor is between 0-5k then more than 25% of the applicant are defaulters.
- ▶ More than 25% of the defaulter applicants paying loan with a interest rate between 9-12%.
- ▶ More than 40% of the applicants whose annual income is between 31k to 58k are defaulters.
- ▶ Approximately 25% of the applicants are defaulters if there debt to income ratio (dti) is between 10-20.
- ▶ Highest number of loan issued was in the year 2011 and in the month december.
- ▶ Applicant belong to grade a is less likely to be a defaulter.
- ▶ Almost 50% of the loan taken by the charged off borrower is for debt\_consolidation.
- ▶ Almost 40% of the charged off applicants opt for 60 months term to pay their loan.



# Numerical -Numerical Bivariate Analysis:

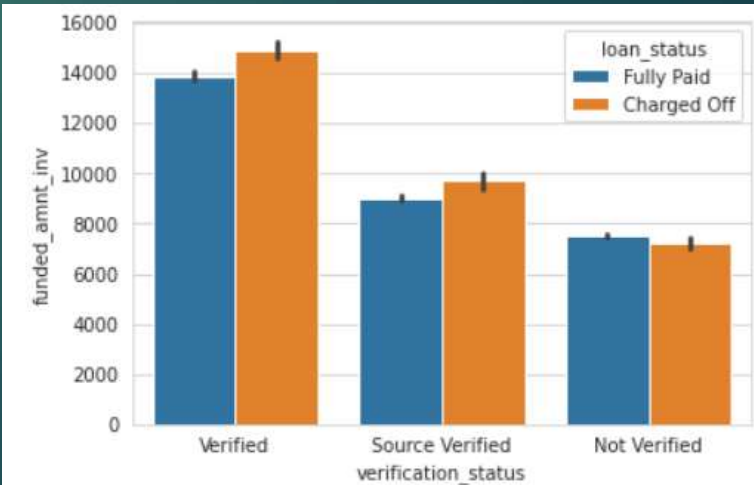
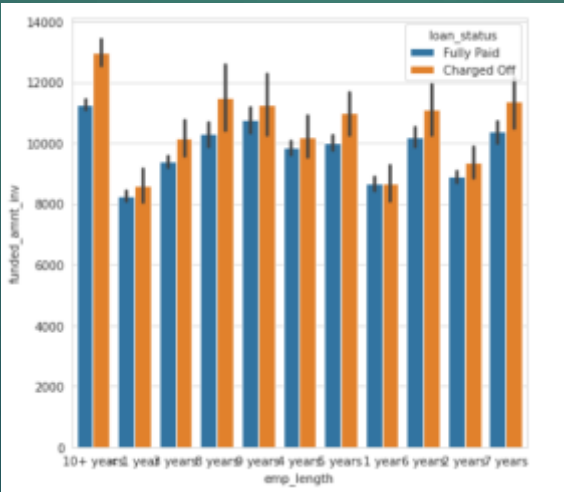
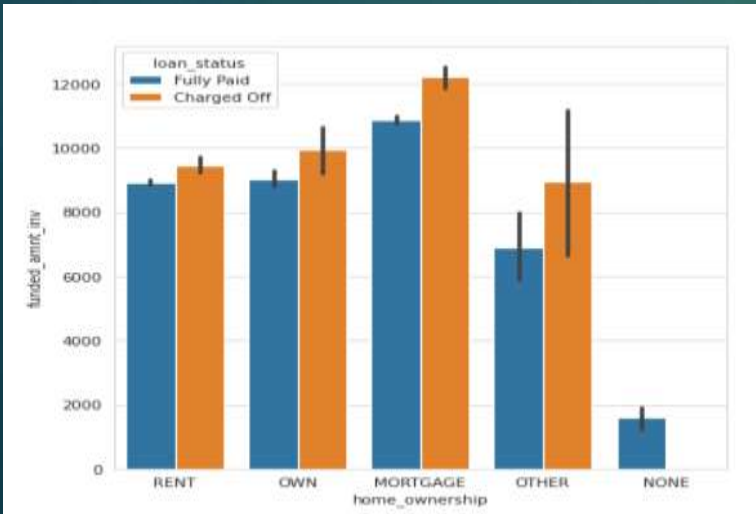
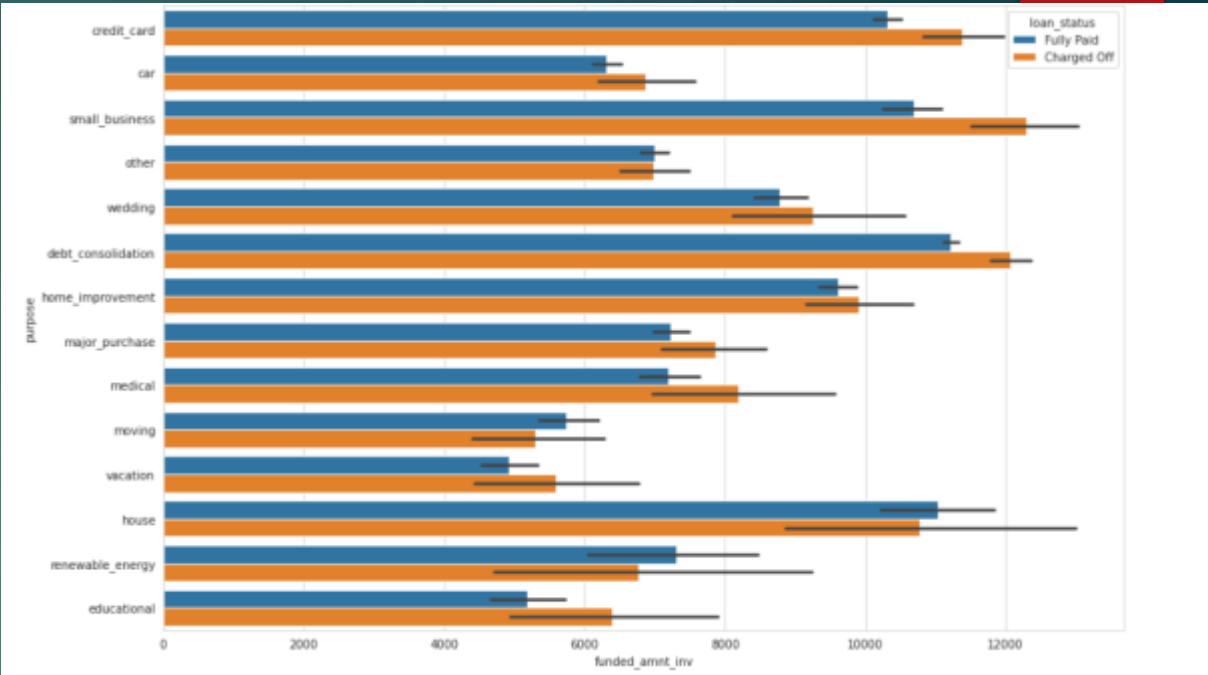
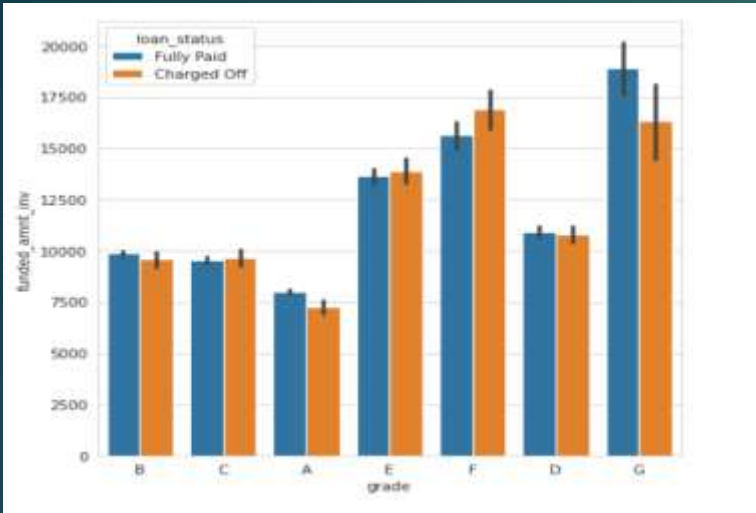


# Conclusion From Heat Map:

- ▶ The columns - "loan amount", "funded amount", "funded amount by inventor", "installment" are highly correlated to each other.
- ▶ The issue\_d\_yy and interest rate are positively correlated to all the variables.
- ▶ The columns - "loan amount", "funded amount", "funded amount by inventor", "installment" are slightly correlated to annual income and interest rate.
- ▶ Dti is negatively correlated to annual income.
- ▶ A cluster is created by issue\_d\_yy, int\_rate and annual\_inc.
- ▶ Pub\_rec\_bankruptcies is negatively related to "loan amount", "funded amount", "funded amount by inventor", "installment" but positively related to interest rate, issue date year, debt to income (dti).

# Numerical -Categorical Bivariate Analysis:

Funded amount by investor v/s categories:

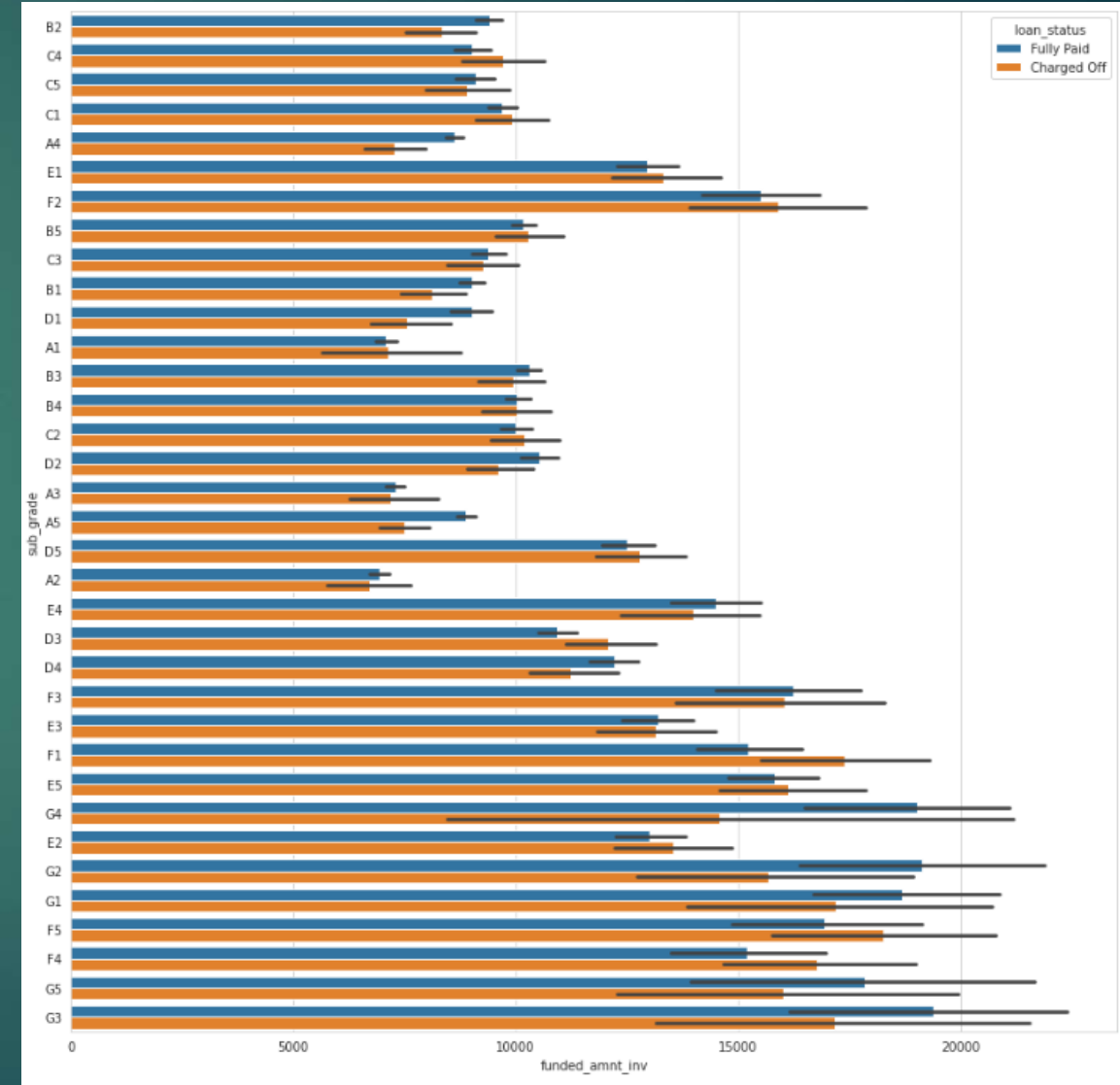


# Numerical -Categorical Bivariate Analysis:

## Funded amount by investor v/s categories:

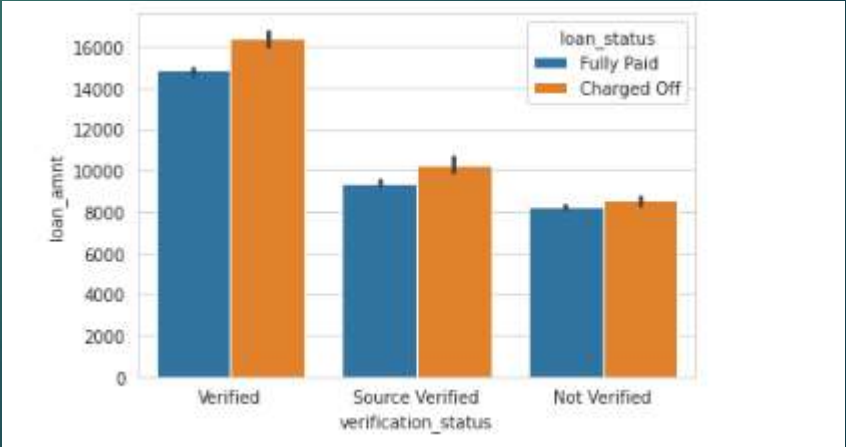
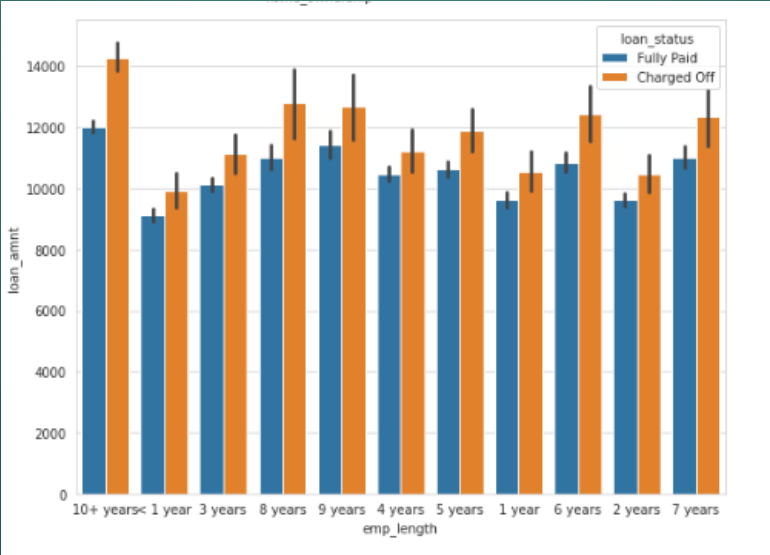
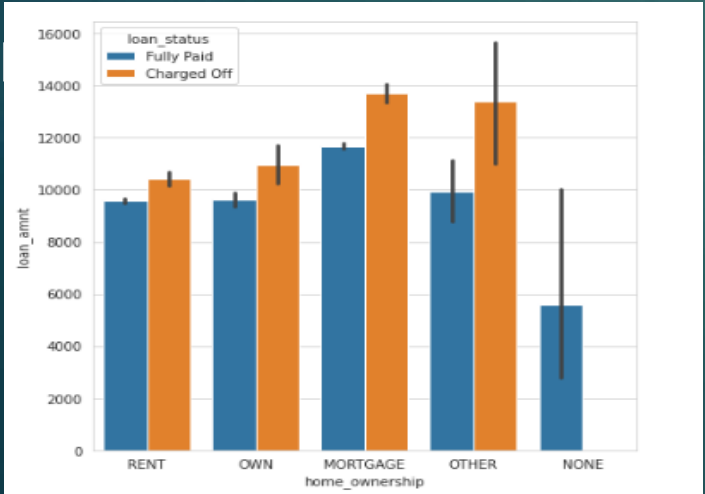
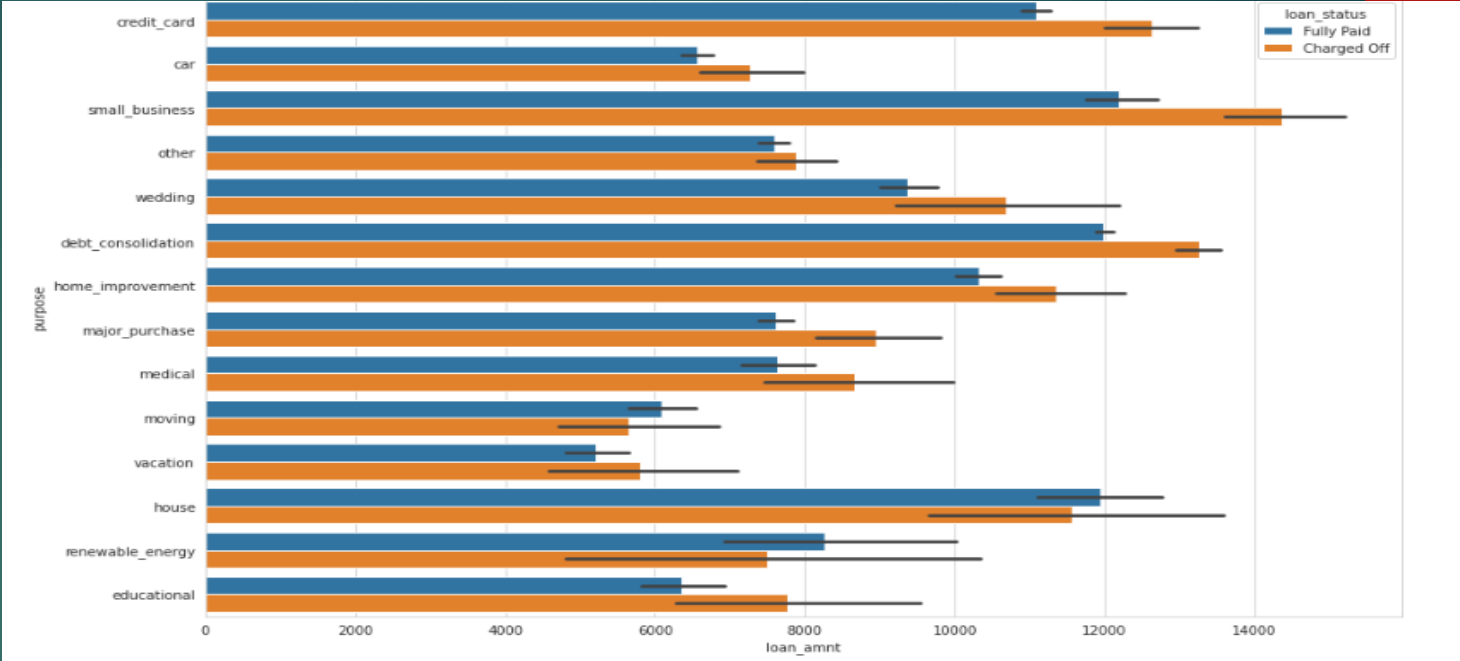
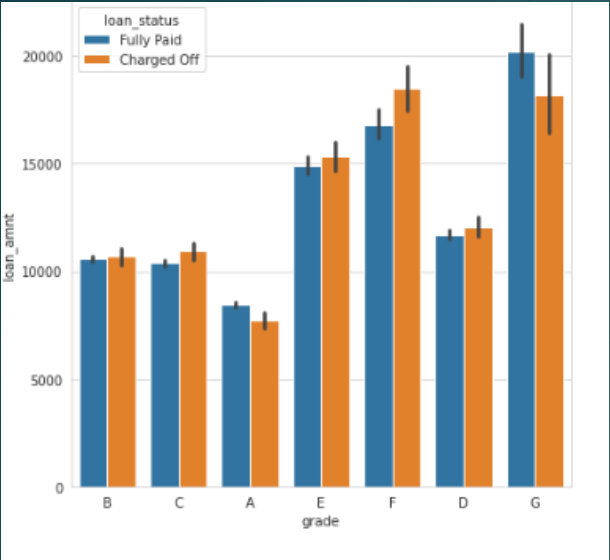
From the above graphs, I can conclude the below points:

- ▶ Investor funded loan around 17k-18k to a borrower belongs to grade F are more likely to be a defaulter. Whereas applicants belongs to grade G and investor funded loan amount between 18k-20k fully pay the bill.
- ▶ Investor funded loan around 15k-20k to a borrower belongs to sub grade f5 are more likely to be a defaulter. Whereas, investor funded loan above 20k to an applicant belongs to sub grade G2 fully pay the loan.
- ▶ Investor funded loan more than 12k to a borrower with a mortgage home ownership can be a defaulter.
- ▶ Borrowers taken loan for the purpose of small business and funded by investor between 10-12k are defaulters.
- ▶ Borrowers with more than 10 years of experience get investor funded loan between 12k and 14k become defaulters.



# Numerical -Categorical Bivariate Analysis:

➤ Loan amount v/s categories:

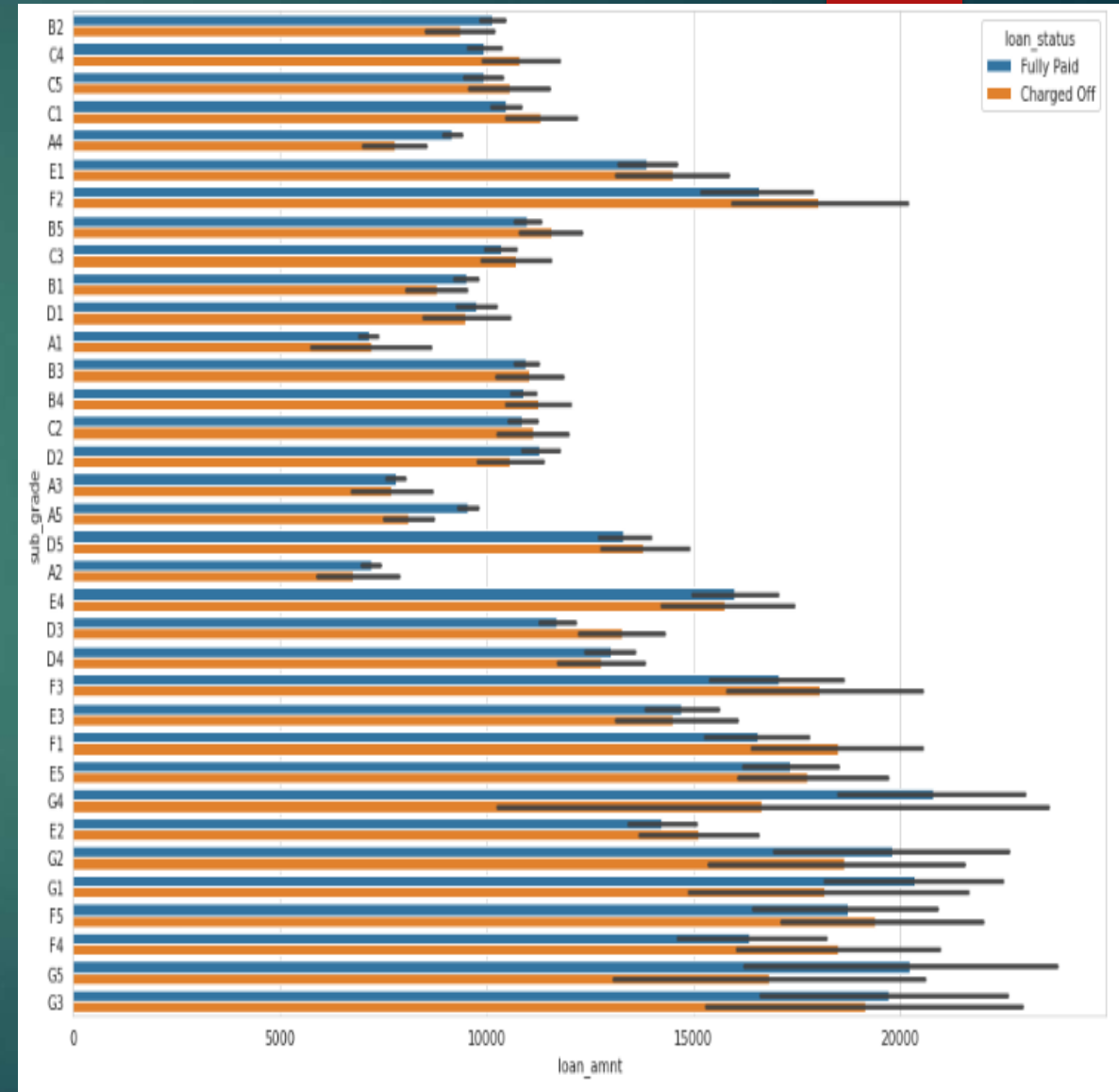


# Numerical -Categorical Bivariate Analysis:

## Loan amount v/s categories:

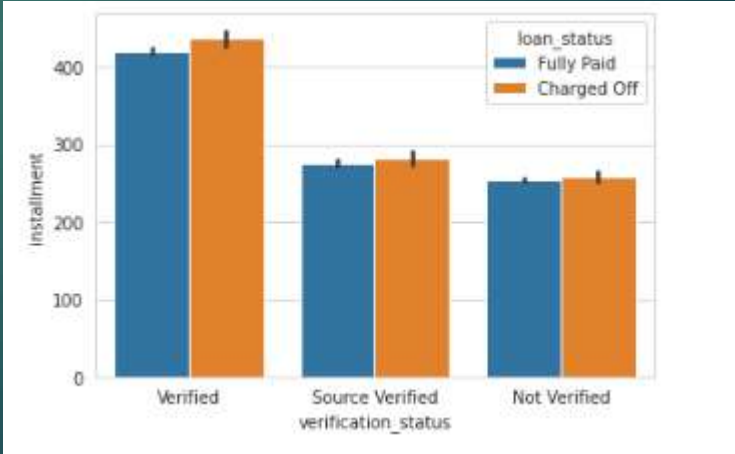
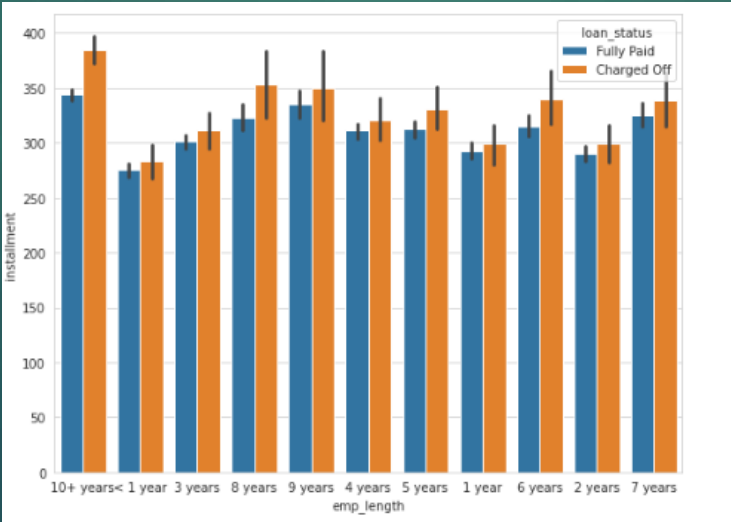
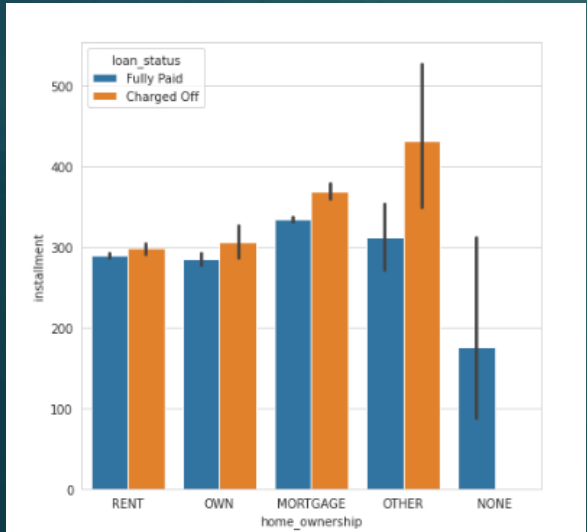
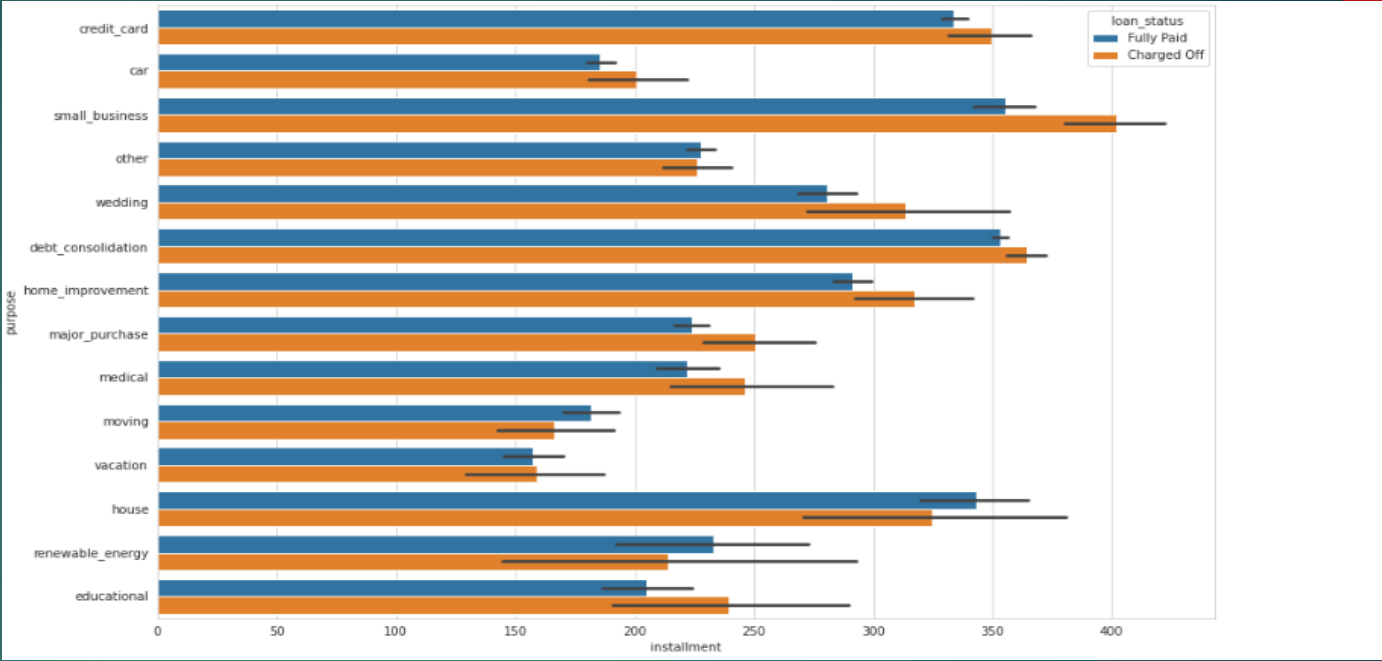
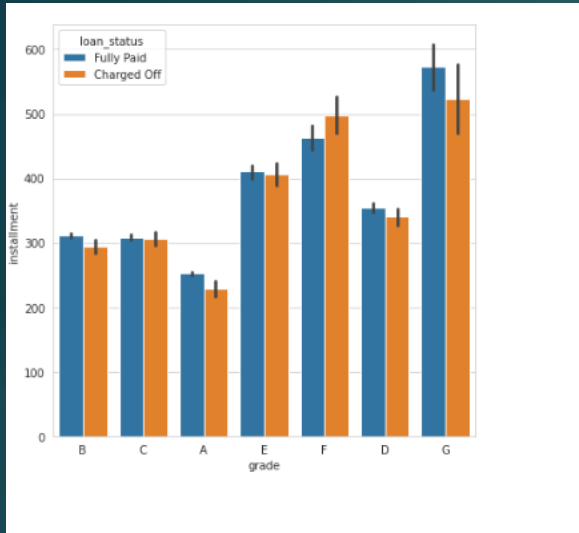
From the above graphs, I can conclude the below points:

- ▶ Borrowers applied for loan amount around 17k-18k and belongs to grade F are more likely to be a defaulter. Whereas applicants belongs to grade G and applied for loan amount above 20k fully pay the bill.
- ▶ Brrowers asked for loan amount between 12-14k to a borrower with a mortgage home ownership can be a defaulter. below 12k there is a 50% probability of applicant can be defaulter.
- ▶ loan applied for the purpose of small business and amount borrowed between 14-16k are defaulters.
- ▶ Borrowers with more than 10 years of experience applied for loan in the range from 12- 14k become defaulters.
- ▶ loan applied above 20k and belongs to sub grade F5 are more likely to be a defaulter. Whereas, applicant loan above 20k belongs to sub grade G2 fully pay the loan.



# Numerical -Categorical Bivariate Analysis:

## ➤ Loan Amount v/s Categories:



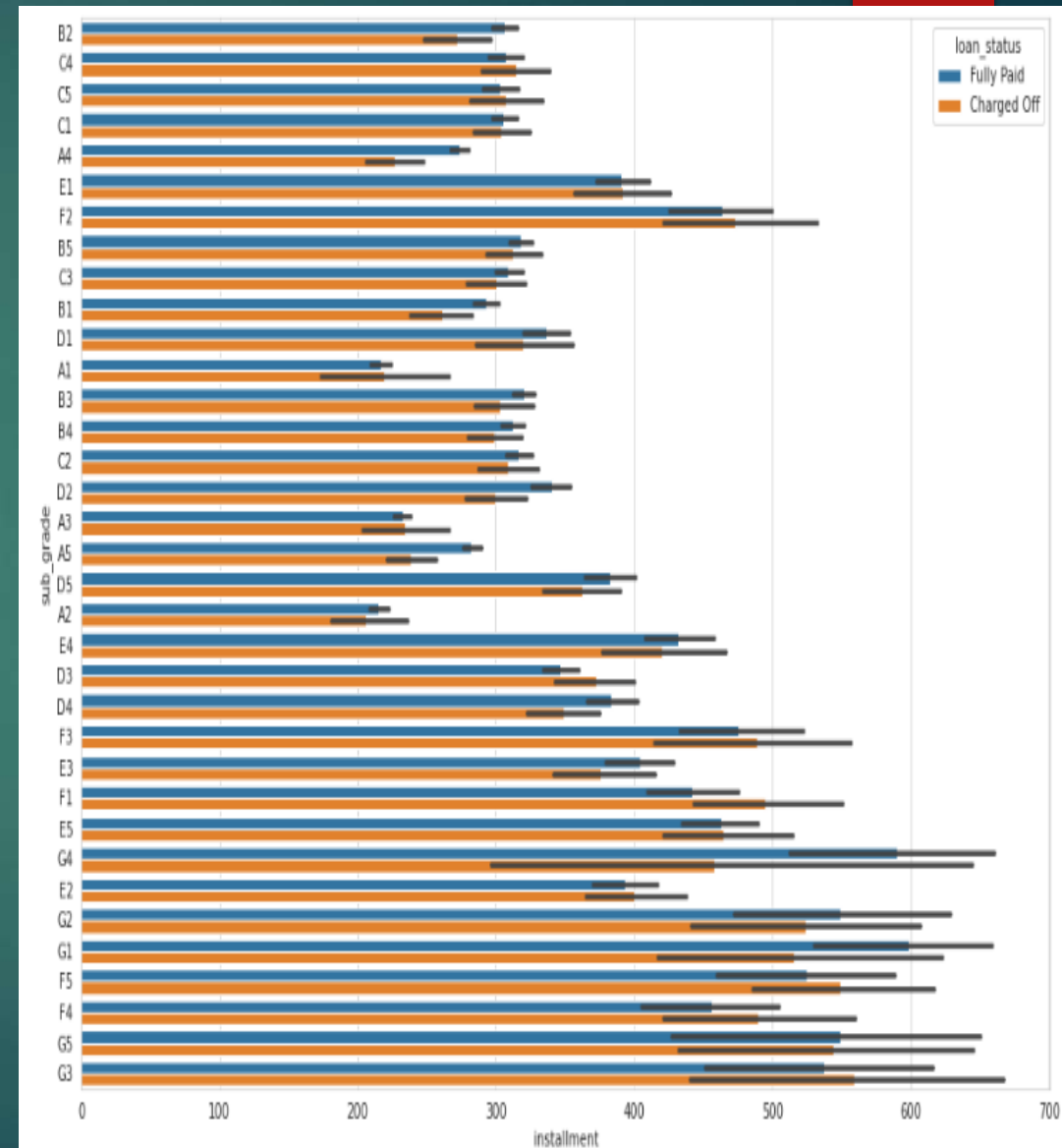


# Numerical -Categorical Bivariate Analysis:

## Loan amount v/s categories:

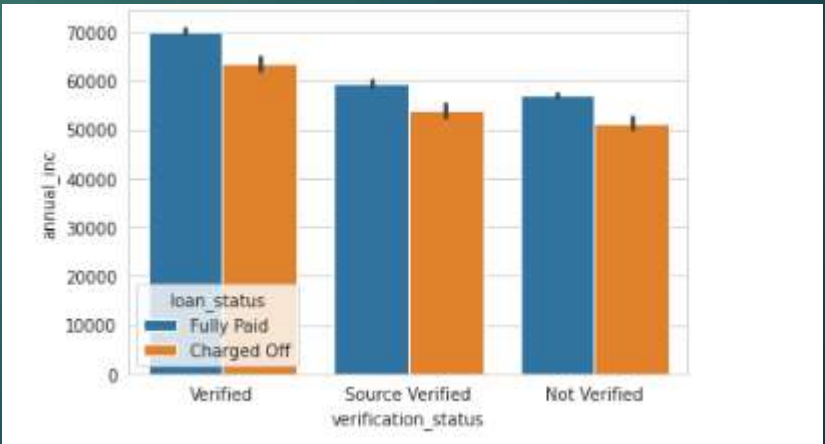
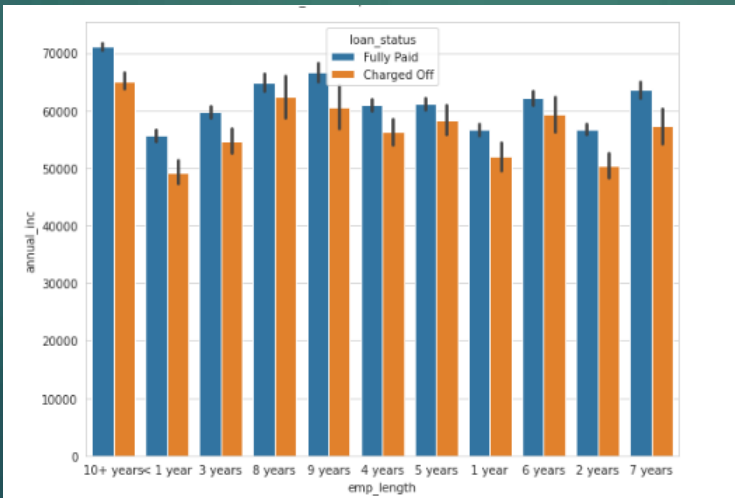
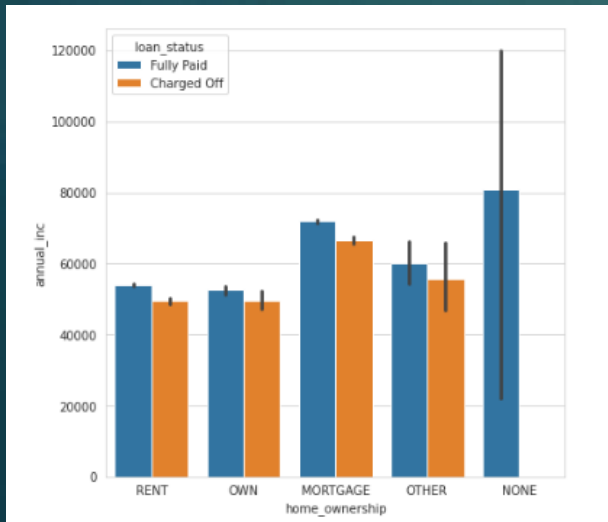
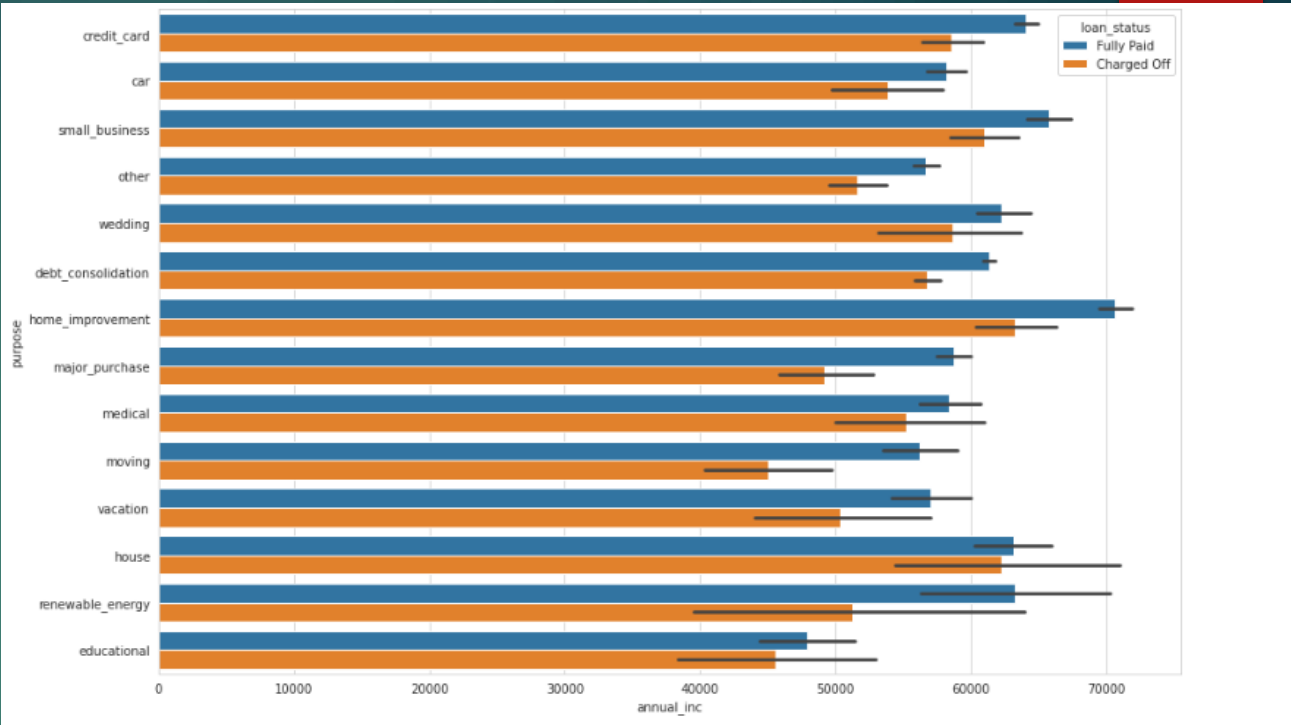
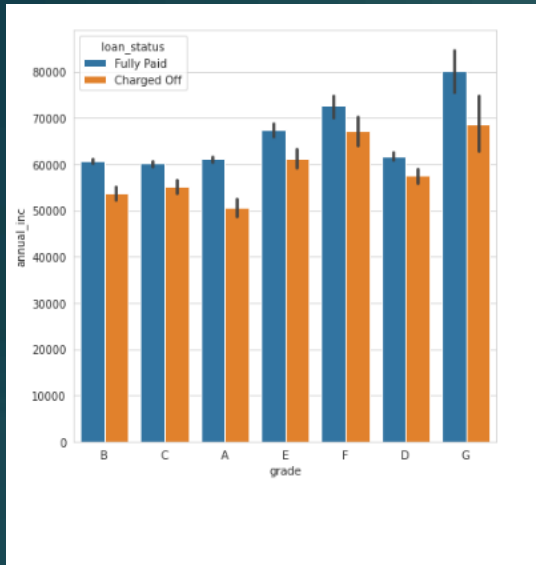
From the above graphs, I can conclude the below points:

- ▶ Applicant belong to grade G and installment above 520 fully pay their loan however, Applicants belong to grade F and installment above 490 can become defaulters.
- ▶ applicants with installment above 400 and with other as home ownership can be a defaulter. However, Borrower with installments between 300-350 and have other as home ownership are more likely to pay the loan.
- ▶ loan asked for the purpose of small business and have monthly installment above 400 are defaulters
- ▶ applicants with more than 10 years of experience and monthly installments between 360-390 can become defaulters. However, Borrowers with installments payment close to 350 are Fully payers.
- ▶ applicants with monthly installment approximately between 570-600 and belongs to sub grade G5 and F5 are more likely to be a defaulter. Whereas, applicant monthly install above 470 belongs to sub grade G4 fully pay the loan.



# Numerical -Categorical Bivariate Analysis:

## Annual Income v/s Categories:



# Numerical -Categorical Bivariate Analysis:

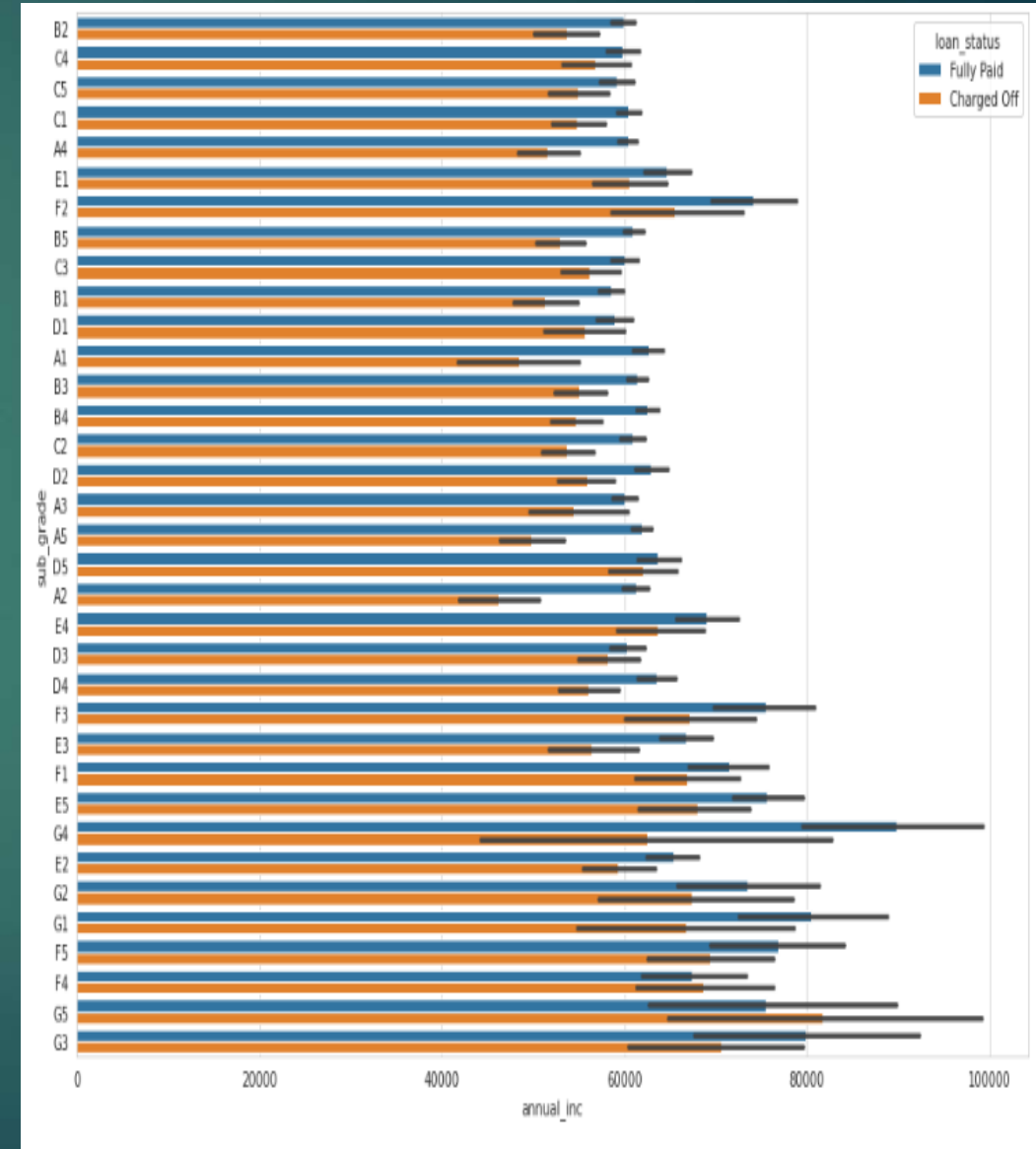
## Annual Income v/s Categories:

From the above graphs, I can conclude the below points:

- ▶ Applicants with annual income above 80k has highest probability to fully pay the loan.
- ▶ Applicant with annual income up to 80k and belong to grade F can become a defaulter.
- ▶ Applicants with annual income below 80k and with Mortgage as home ownership can be a defaulter. However, above 80k and Mortgage as home ownership are more likely to pay the loan.
- ▶ loan asked for the purpose of Home improvement and with annual salary approximately below 79k can become defaulters.
- ▶ Applicants with more than 10 years of experience and annual income below 72k can become defaulters.
- ▶ applicants with annual income above 1L and belongs to sub grade G4 more likely to be a defaulter. Whereas, applicant with annual income above 1L and belong to same group are more likely to pay the loan.

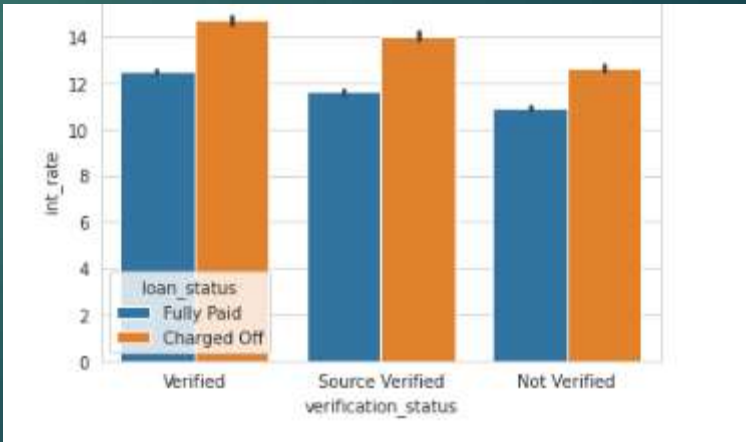
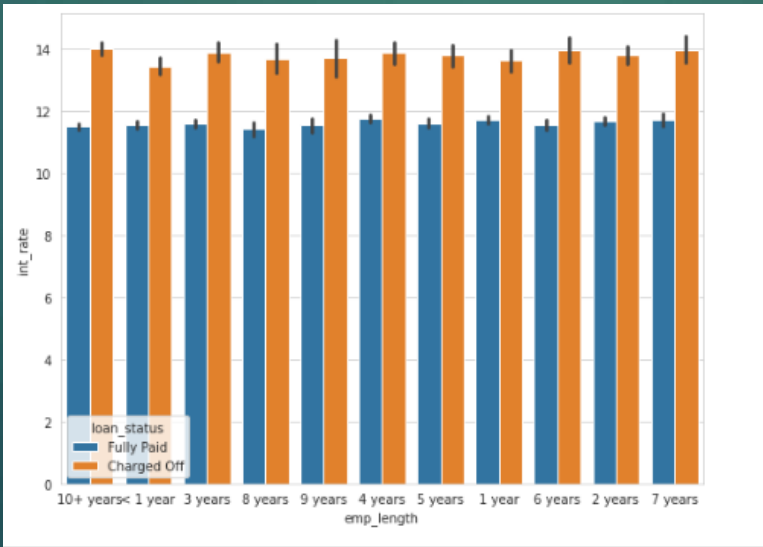
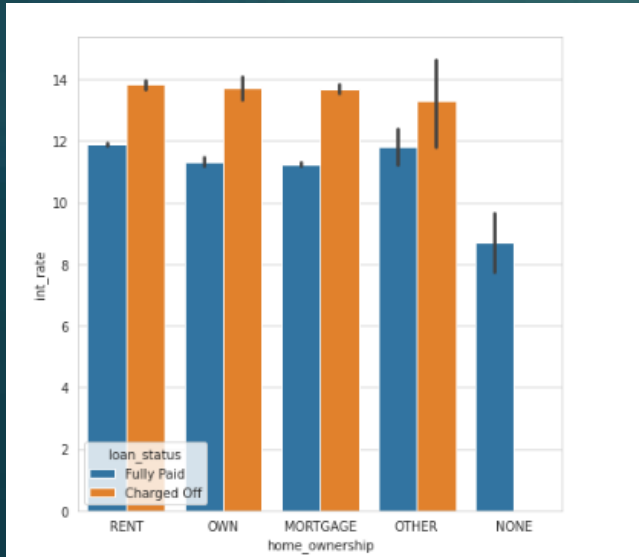
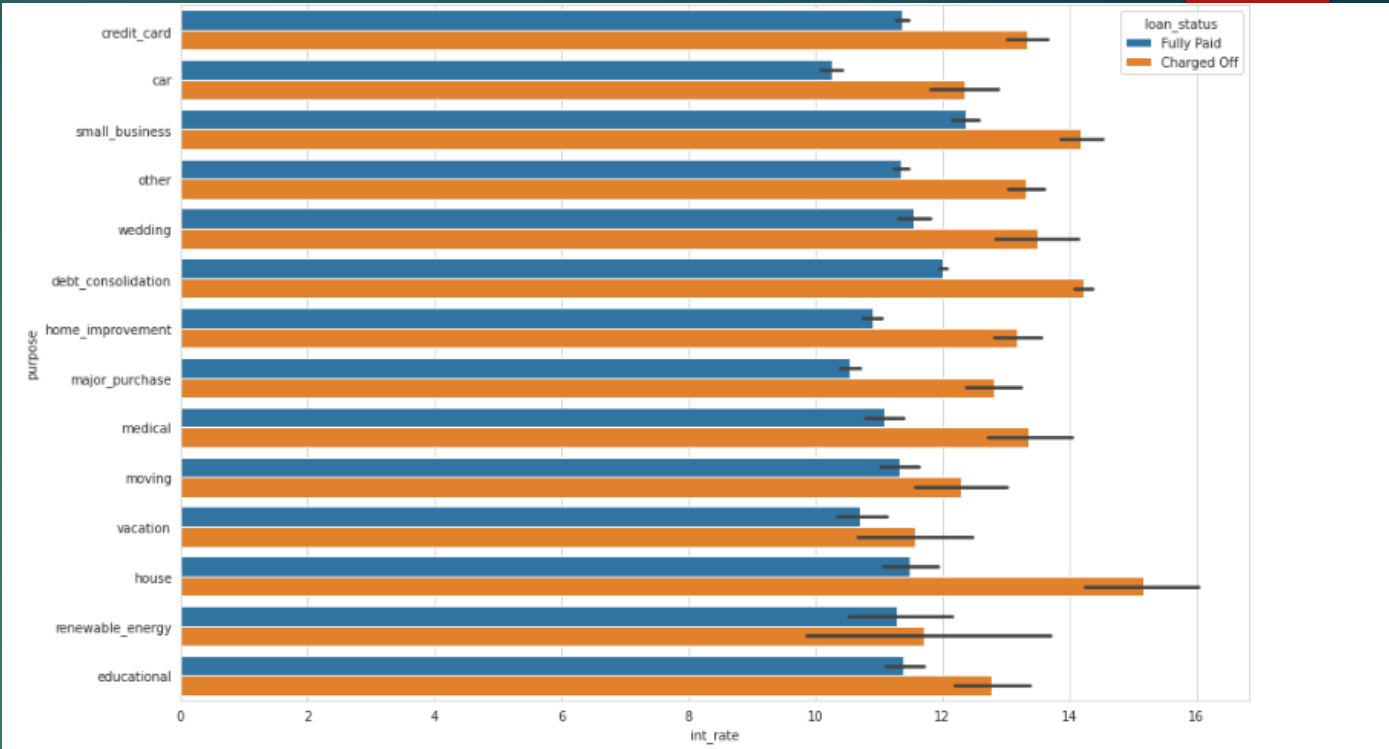
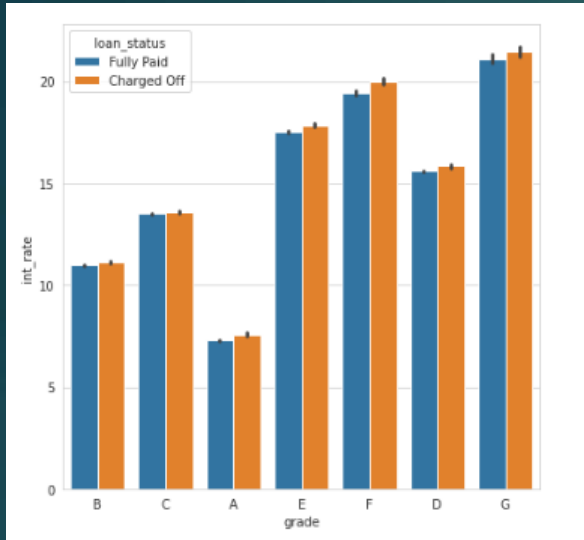
Recommendation:

- ▶ LC should give more loans to the applicants whose annual salary is above 80k.



# Numerical -Categorical Bivariate Analysis:

➤ Interest Rate v/s Categories:

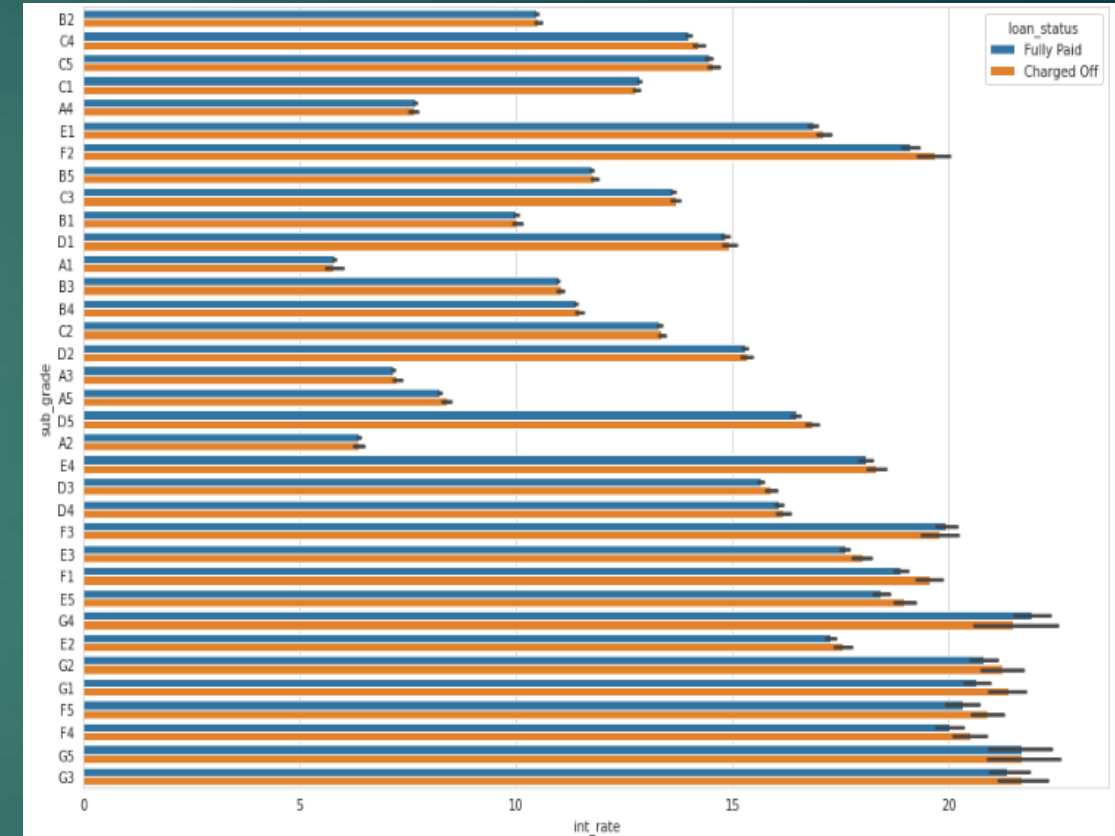


# Numerical -Categorical Bivariate Analysis:

## Interest rate v/s Categories:

From the above graphs, I can conclude the below points:

- ▶ Applicants with interest rate above 12% are very much likely to become a defaulter.
- ▶ Applicants with any employment experience paying loan with interest rate above or equal to 12% become a defaulter.
- ▶ Applicants applied loan for any purpose if paying loan with interest rate above or equal to 12% become a defaulter.
- ▶ Applicants with verification status as non verified/source verified paying loan with interest rate above or equal to 12% become a defaulter. However, for verified status if the rate is higher than 12.5% then the applicant become a defaulter.
- ▶ There is no evident classification between fully paid and charged off applicants for the grade and sub-grade variable with respect to the interest rate.

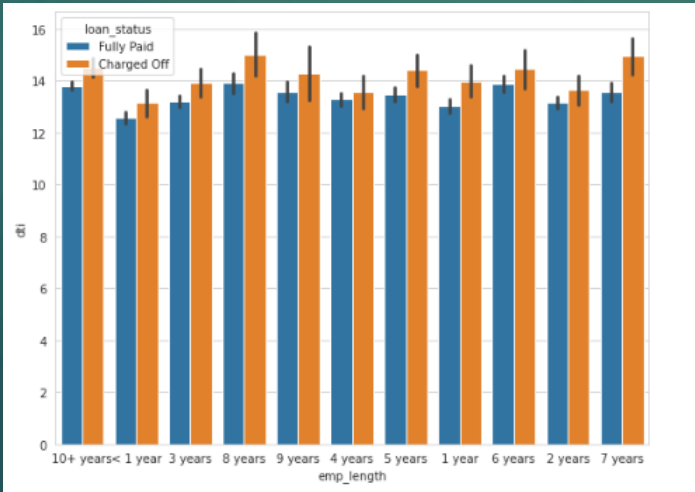
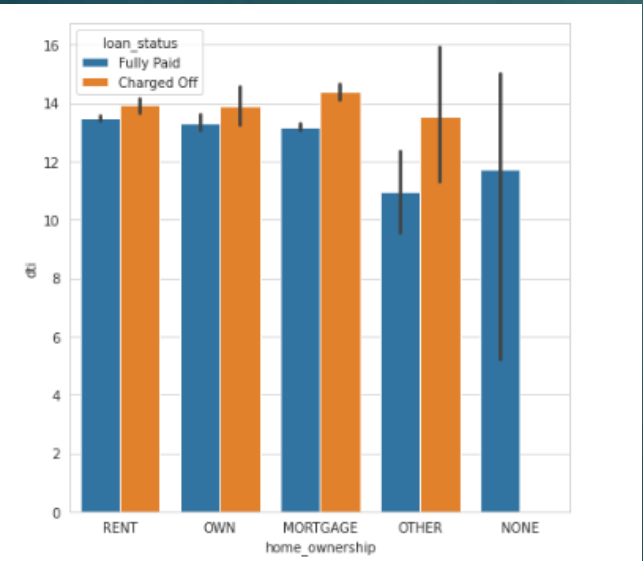
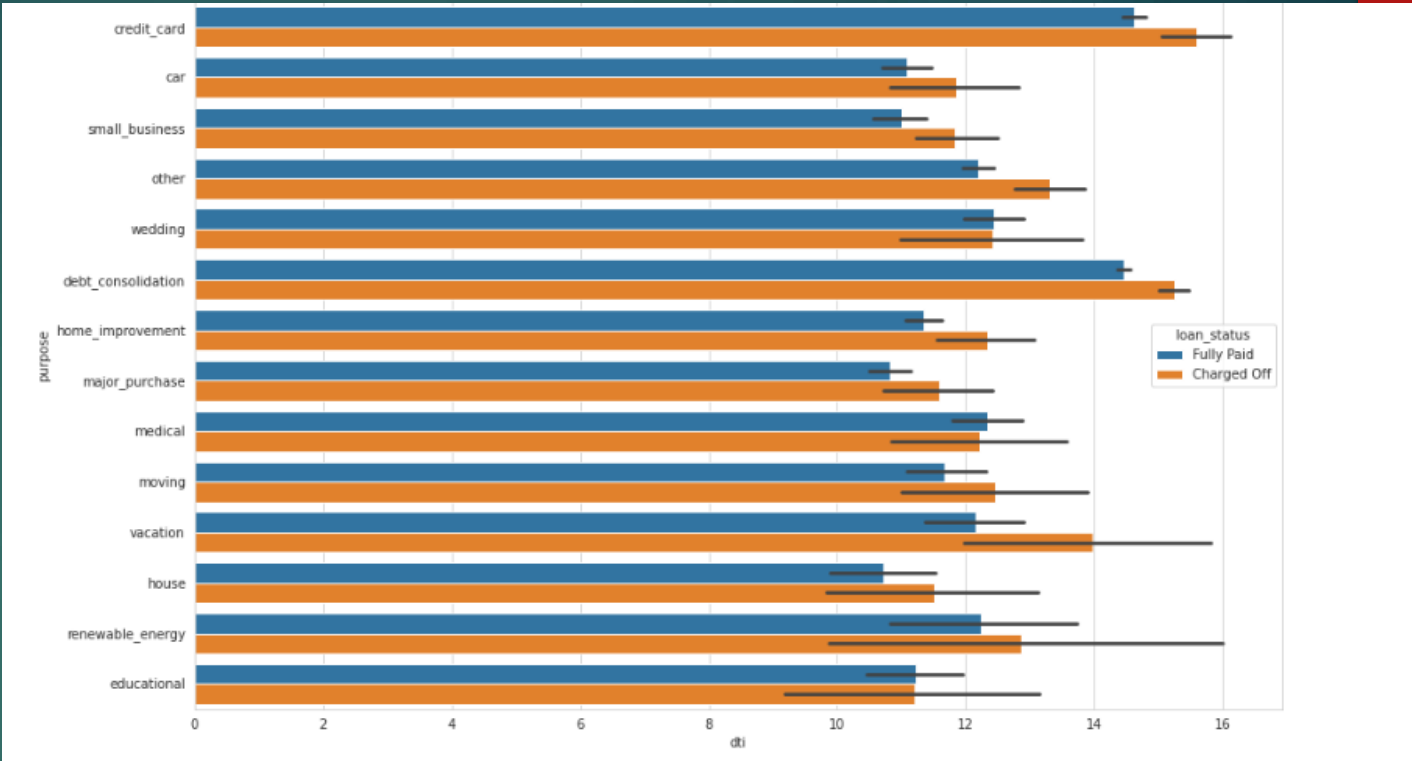
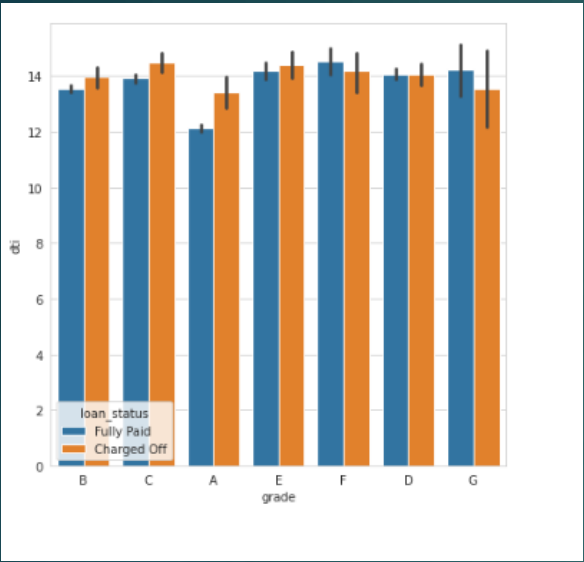


## Recommendation:

- ▶ LC should decrease the interest rate in order to decrease the probability of defaulters.

# Numerical -Categorical Bivariate Analysis:

➤ Dti v/s Categories:

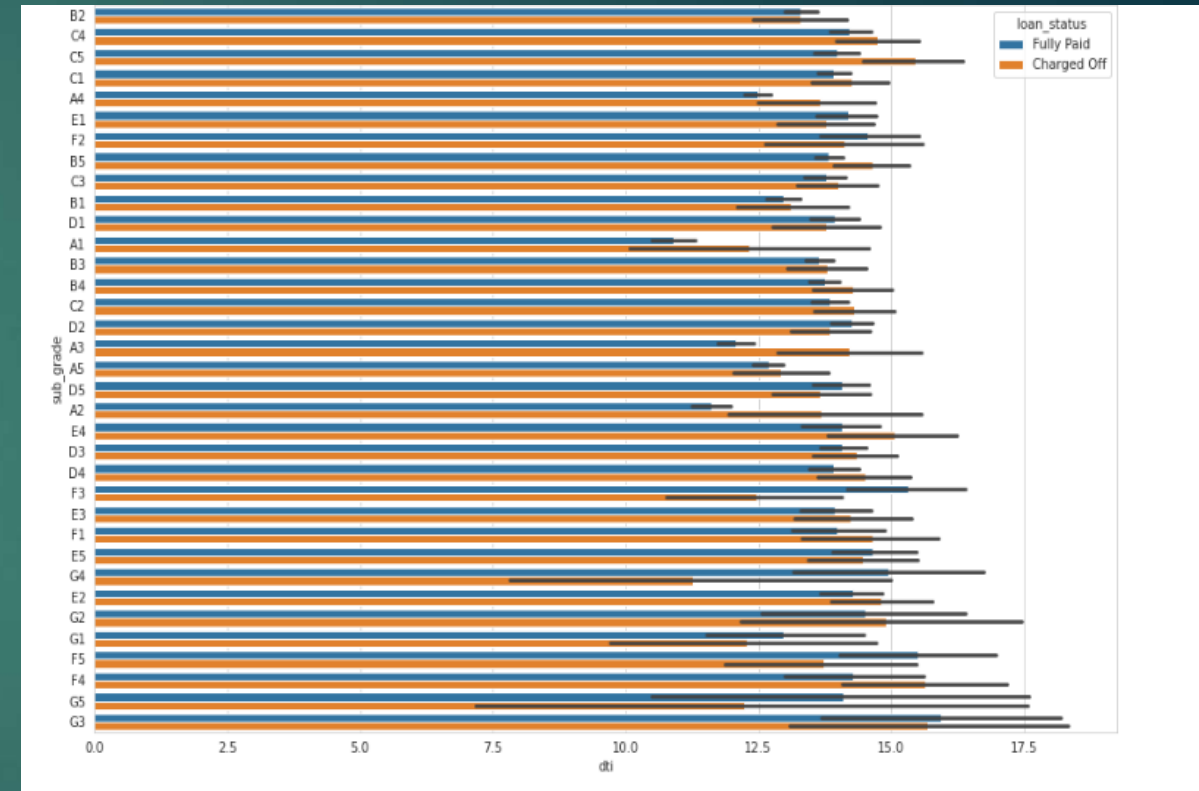


# Numerical -Categorical Bivariate Analysis:

## Dti v/s Categories:

From the above graphs, I can conclude the below points:

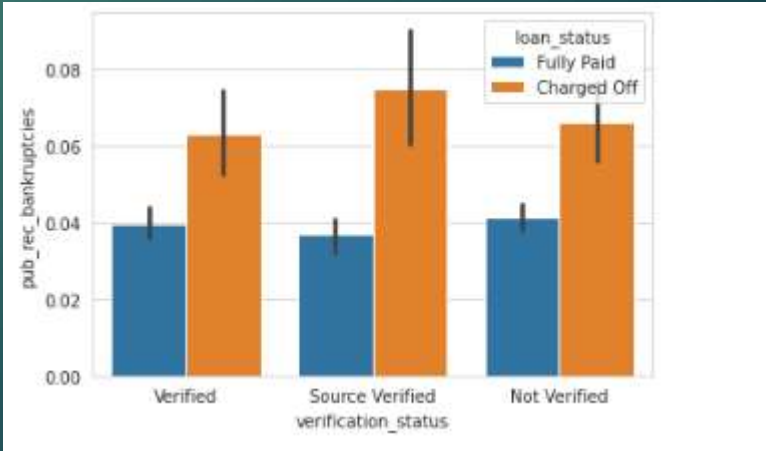
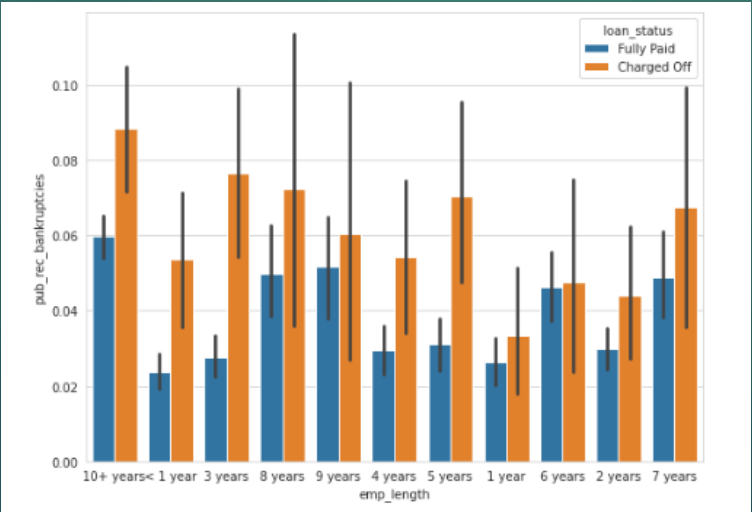
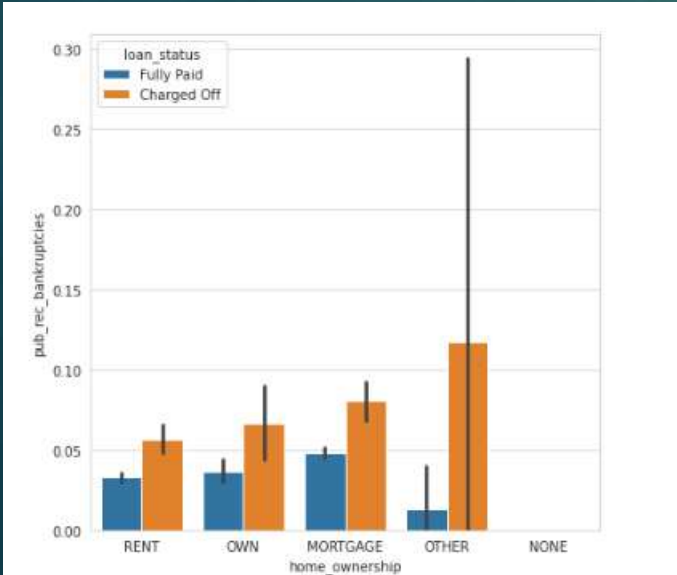
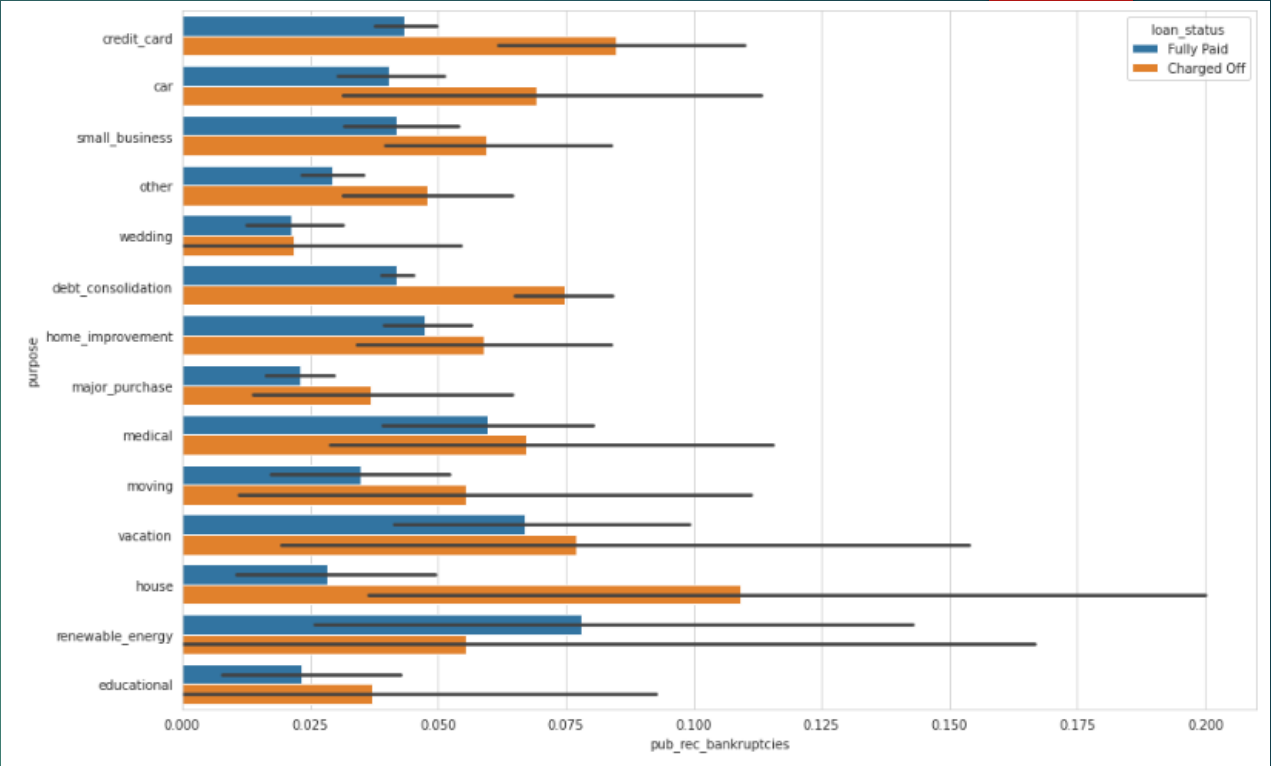
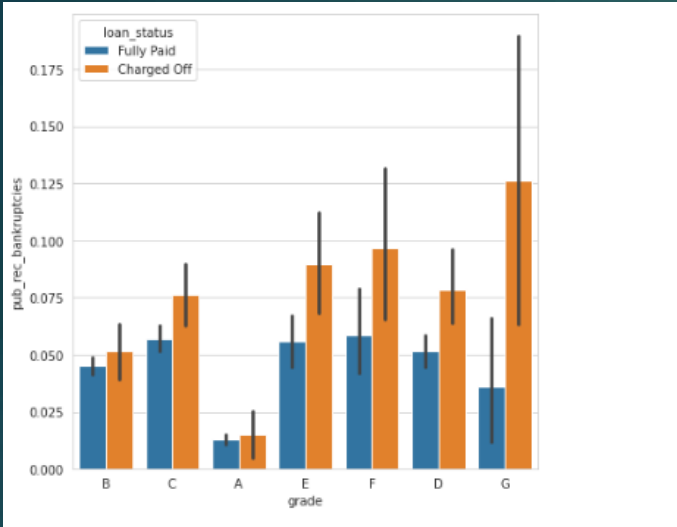
- ▶ Applicants with any employment experience dti above or equal to 14 become a defaulter.
- ▶ Applicants applied loan for purpose other than credit card and debt consolidation have dti above 12.5 have a risk of becoming a defaulter.
- ▶ Applicants with verification status as non verified/source verified have dti above or equal to 12.5 become a defaulter. However, for verified status if dti is higher than 14 then the applicant can become a defaulter.
- ▶ If the dti of a applicant is above 14.00 and grade is C then the applicant can become a defaulter. However, for Grade F and G applicant with dti higher than 14 can fully pay the loan.
- ▶ For all sub grade of grade C the probability of applicant becoming defaulter is higher. But for sub grade F3 and dti greater than 13.00 the probability of applicant fully pay the loan is higher.





# Numerical -Categorical Bivariate Analysis:

➤ Public Record of Bankruptcies v/s Categories:

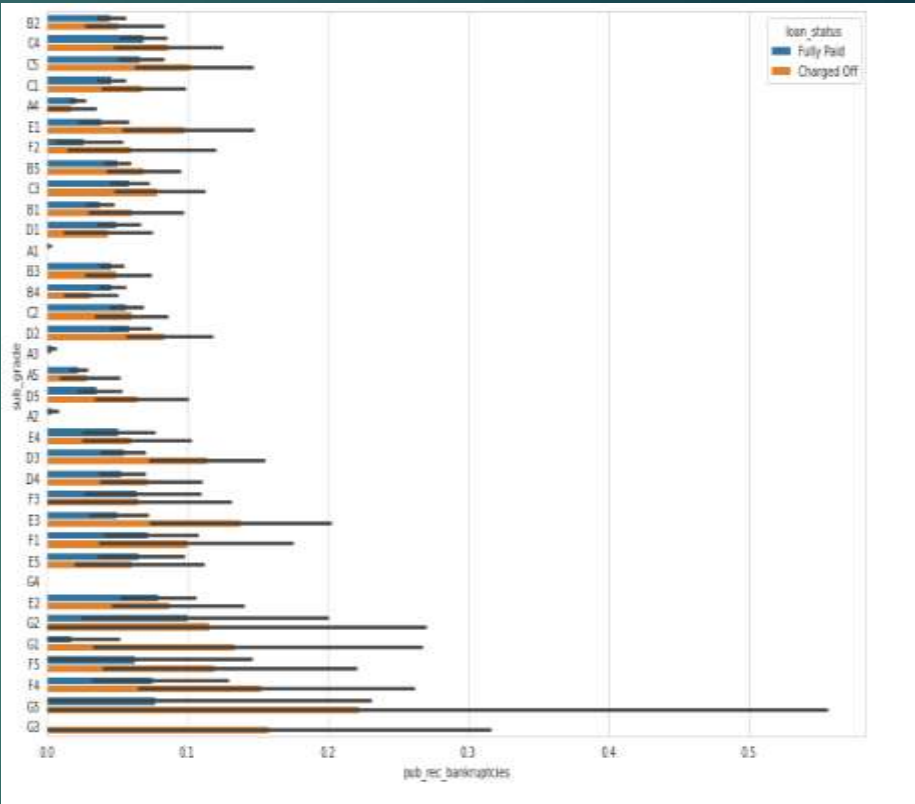


# Numerical -Categorical Bivariate Analysis:

## Public Record of Bankruptcies v/s Categories:

From the above graphs, I can conclude the below points:

- ▶ Applicants with greater than 0 are more to prone to become defaulter.



# Conclusion from Numerical-Categorical Bivariate Analysis

- ▶ Investor funded loan or loan applied for around 17k-18k to a borrower belongs to grade F are more likely to be a defaulter.
- ▶ Investor funded loan around 15k-20k to a borrower belongs to sub grade F5 are more likely to be a defaulter.
- ▶ Investor funded loan more than 12k to a borrower with a mortgage home ownership can be a defaulter.
- ▶ Borrowers taken loan for the purpose of small business and funded by investor between 10-12k or loan applied between 14-16k are defaulters.
- ▶ Borrowers with more than 10 years of experience applied for loan in the range from 12- 14k become defaulters.
- ▶ loan applied above 20k and belongs to sub grade F5 are more likely to be a defaulter.
- ▶ Applicants belong to grade F and installment above 490 can become defaulters.
- ▶ applicants with installment above 400 and with other as home ownership can be a defaulter.
- ▶ loan asked for the purpose of small business and have monthly installment above 400 are defaulters.
- ▶ applicants with monthly installment approximately between 570-600 and belongs to sub grade G5 and F5 are more likely to be a defaulter.
- ▶ Applicant with annual income upto 80k and belong to grade F can become a defaulter.
- ▶ Applicants with annual income below 80k and with Mortgage as home ownership can be a defaulter.
- ▶ Applicants with annual income above 80k has highest probability to fully pay the loan.
- ▶ loan asked for the purpose of Home improvement and with annual salary approximately below 79k can become defaulters.
- ▶ Applicants with more than 10 years of experience and annual income below 72k can become defaulters.

# Conclusion from Numerical-Categorical Bivariate Analysis

- ▶ Applicants with interest rate above 12% are very much likely to become a defaulter.
- ▶ Applicants with any employment experience paying loan with interest rate above or equal to 12% become a defaulter.
- ▶ Applicants applied loan for any purpose if paying loan with interest rate above or equal to 12% become a defaulter.
- ▶ Applicants with verification status as non verified/source verified paying loan with interest rate above or equal to 12% become a defaulter. However, for verified status if the rate is higher than 12.5% then the applicant become a defaulter.
- ▶ Applicants with any employment experience dti above or equal to 14 become a defaulter.
- ▶ Applicants applied loan for purpose other than credit card and debt\_ consolidation have dti above 12.5 have a risk of becoming a defaulter.
- ▶ Applicants with verification status as non verified/source verified have dti above or equal to 12.5 become a defaulter.
- ▶ If the dti of a applicant is above 14.00 and grade is C then the applicant can become a defaulter. However, for Grade F and G applicant with dti higher than 14 can fully pay the loan.
- ▶ Applicants with greater than 0 are more to prone to become defaulter.

# Important Driver Variables:

- ▶ Debt to income(dti) ratio more than 12.5
- ▶ Grade of a applicant especially if the grade is F and C if the dti is high.
- ▶ Sub-grade of a applicant especially f5.
- ▶ Annual income less than 80k.
- ▶ House ownership as mortgage.
- ▶ Interest rate mire than 12%
- ▶ Number of bankruptcies more than 0.
- ▶ Purpose of loan are - debt\_consolidation, small business.

Thank you